Developing a novel spatial memory task for early Alzheimer's diagnosis with immersive virtual reality and eye-tracking

Luke Emrich-Mills

A thesis submitted for the degree of
Doctor of Philosophy
in Clinical Cognitive Neuroscience

Supervised by
Professor Neil Burgess | Institute of Cognitive Neuroscience
Professor John King | Clinical, Educational and Health Psychology

Institute of Cognitive Neuroscience
University College London

December 2023

Declaration of Authorship

I, Luke Emrich-Mills, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Abstract

Alzheimer's disease (AD) involves pathophysiological damage in the medial temporal lobe (MTL), an area of the brain implicated in spatial processing. Consequently, testing spatial cognition has become a promising approach for identifying early behavioural changes in AD compared to healthy aging. Immersive virtual reality (iVR) is a suitable medium for administering spatial tasks due to its naturalistic spatial interaction and controlled stimulus presentation, with an iVR path integration test already demonstrating impressive diagnostic potential for early AD. Additionally, modern head-mounted virtual reality displays come equipped with integrated eye tracking, permitting combined exploration of AD-related eye movement changes alongside spatial testing. In this thesis, I describe the development of a novel iVR spatial memory task with concomitant eye-tracking for detecting ageing- and AD-related behavioural and gaze fixation changes. The task was based on a classic viewpoint-shifting paradigm that was previously untested in older or memory-impaired participants. The design, development, and feasibility of the task are reported prior to results on the diagnostic potential of eye movement and memory metrics for discriminating participants at high risk of AD from lower-risk individuals, with a secondary focus on healthy age-related differences. Participants with Mild Cognitive Impairment and presence of AD biomarkers (MCI+) were compared to those without biomarkers (MCI-) and to healthy, age-matched adults. Healthy younger participants were also recruited to examine aging effects. Group differences were observed in how, where and when participants viewed spatial stimuli during memory encoding and retrieval. These effects extended beyond differences in task performance, were dependent on spatial condition, and outperformed traditional neuropsychological tests in identifying MCI+ participants from controls. Findings provide early support for a spatial eye-tracking task as a promising addition to future diagnosis of AD.

Impact Statement

As the leading cause of dementia, Alzheimer's disease (AD) poses a major global health challenge, with the first disease-modifying treatments only recently approved for clinical application. Earlier detection of the disease has been proposed as a contributing factor to the success of future pharmacological treatments. This thesis presents primary evidence that a novel spatial memory test, augmented with eye-tracking, could improve early diagnosis of AD beyond the capabilities of existing clinical tests. This builds upon previous work demonstrating the diagnostic utility of spatial tests hosted in immersive virtual reality.

Any impact on AD diagnosis derived from this work would not be immediate. Larger-scale diagnostic trials are required to establish the accuracy of any potential new test, including studies that validate immersive technology for use in clinical practice. However, an overarching aim of this line of research is to inform and transform diagnostic practices on a wide scale, potentially influencing regional, national, and even international approaches.

Aside from clinical diagnosis, the findings in this thesis could influence future research either in this field or other disciplines employing similar methodologies. For example, employing eye tracking as a measure of memory retrieval was a novel approach for the spatial paradigm employed in this thesis, setting a precedent for subsequent research into spatial cognition using this technique. Indeed, differences in viewing patterns were found between healthy younger and older adults, which may contribute to our understanding of memory processing in ageing. Alternatively, the findings related to the use of virtual 'teleportation' could support work that makes use of virtual reality technology more broadly.

Acknowledgements

I would like to thank my supervisors, Professors Neil Burgess and John King, for providing me with the opportunity and support to complete this doctorate. It has been a huge privilege to work in this field and learn from such impressive men. They have been exemplary mentors, especially during the more difficult moments.

I would also like to thank Professor Dennis Chan, who has been a frequent source of inspiration before and during this PhD. Without him, I would have had no way of reaching patients, and the potential impact of this work may have been considerably reduced. His support when deciding my next steps was also above and beyond reasonable expectations.

Thanks also to Professor Joshua Stott, the final member of my thesis committee, who's supportive and encouraging words were always welcome during our occasional meetings.

Thanks to the Medical Research Council for funding this work, and all the staff at the UCL-Birkbeck MRC Doctoral Training Programme for providing the support and opportunities to expand my horizons and potentially contribute to something important. Thanks also to HTC for gifting one of the Vive Pro Eye headsets used in the project.

A huge thank you to Dr Andrea Castegnaro, who set the precedent for this project with his own PhD, and trained me to build virtual environments from scratch. He gave a crash course in Unity development early on, which allowed me to see my ideas come to (virtual) reality. Beyond this PhD, I am certain we will continue to collaborate in this exciting area of work.

Thanks to members of Neil's Space & Memory lab for contributing to a warm and friendly working environment. Special thanks to Kasia, for making those long train journeys down to Haywards Heath much more tolerable.

The Research and Development team at University Hospitals Sussex also

deserve my thanks for the significant time they put into remedying delays and setbacks. Particular mention goes to Anna Koniotes, who dedicated hours of unpaid time to the clinical study, and to Keely Stewart, who worked tirelessly to implement the administrative overhaul required to get the project back off the ground.

A special mention and thanks to my sister and graphic designer Lizzie Emrich-Mills, who created the original illustration seen in Figure 1.1 that I have adapted and used throughout various presentations of this work.

Most importantly, this research would not have been possible without willing participants. I would therefore like to dedicate this thesis to all the patients and control volunteers who gave their time and energy in the hope that this project would make a difference to others in the future.

Finally, I would like to thank my partner, Beth, who has done more to support me during this process than anyone else. She will be very happy it's over.

Overview

The aim of this thesis was to develop a new spatial memory task in immersive virtual reality as a potential contributor to early diagnosis of Alzheimer's disease (AD). The task required detection of a moved object under different subject and scene rotations. This allowed manipulation of egocentric (subject-to-object), allocentric (object-to-object), and self-motion (walking) contributions to spatial memory, including movement by instant transposition ('teleporting') as a comparator to walking. Memory monitoring abilities were assessed by a simple confidence rating, and gaze fixation patterns were analysed, extending measurement of memory-related cognition beyond just task performance.

Chapter 1 introduces the background to this line of research, providing a rationale for targeting early AD using spatial cognition and eye movements. A review of clinical Alzheimer's diagnostics is outlined to set the thesis aims in context, before describing relevant cognitive neuroscience findings in the fields of spatial cognition, memory monitoring, and eye movements.

Based on this background, I predicted that participants with early AD would show impairments in self-motion and allocentric processing during spatial memory, measured by task performance and eye movements on key visual areas of interest. I also expected memory monitoring accuracy to be lower in AD, measured by a decoupling of confidence ratings from task performance. Crucially, I expected task measures to outperform traditional neuropsychological tests in predicting presence of AD.

Before delving into results, Chapters 2 and 3 describe the development of the task and relevant methodological details. Chapter 2 outlines the general approach for the whole thesis, whereas Chapter 3 details a feasibility study to calibrate and refine the task before testing on patients with mild cognitive

impairment (MCI).

The remainder of the thesis provides a series of analyses comparing task measures across different participant groups. Chapter 4 starts by introducing the participant sample, which consisted of healthy younger and older adults, as well as a group of participants with MCI split by results of cerebrospinal fluid biomarkers of Alzheimer's pathology: MCI participants with positive biomarkers were recruited as an early AD group.

Chapter 4 then describes more detailed hypotheses before covering a broad range of group- and condition-level differences across numerous task measures. This is the most results-dense chapter, and accordingly, frequent summaries will be provided to aid understanding. Overall, I found that spatial memory results were unexpectedly similar between MCI groups, but that eye movements showed greatest differences between biomarker-positive and biomarker-negative participants, although not in the hypothesised condition.

Chapter 5 extends key findings by investigating time-dependent changes in gaze behaviour. A bespoke methodology is described, including two alternative approaches to statistically comparing time-series data between groups. Findings in this chapter revealed that further ageing-related and AD-related differences were present when examining gaze measurements at a more granular level, but results otherwise corroborated those from the previous chapter.

Perhaps the most important analysis, Chapter 6 presents findings showing that eye movement measures in the simplest egocentric condition discriminated biomarker groups with superior accuracy than traditional neuropsychological metrics. However, a combination of conventional tests and eye movements showed the greatest potential to classify biomarker-positive participants overall.

To complete the main body of the thesis, Chapter 7 presents a summary of all key findings throughout the thesis, discussing them in the context of previous research, Alzheimer's diagnostics, and methodological limitations.

An additional methodology was developed for clustering eye movement patterns during object-location encoding. This analysis did not yield ageing- or AD-related results due to methodological limitations, but is provided in Appendix A as a foundation for future work.

Contents

List of Figures

List of Tables

List of Abbreviations

AD	Alzheimer's Disease
MTL	Medial Temporal Lobe
iVR	immersive Virtual Reality
MCI	Mild Cognitive Impairment
MCI+/-	MCI with/without presence of AD biomarkers
Aβ	Amyloid-beta
p-Tau	phosphorylated Tau
CSF	CerebroSpinal Fluid
4MT	4 Mountains Test
PIT	Path Integration Test
AOIs	Areas of Interest for eye fixation analysis
VR-IDT	Virtual Reality adaptation of fixation identification algorithm 'Identification by Dispersion Threshold'
NART	National Adult Reading Test
ACE-III	Addenbrooke's Cognitive Examination, Third Edition
FCSRT	Free and Cued Selective Reminding Test
Rey	Rey-Osterrieth Complex Figure Test
TMT B	Trail-Making Test part B
DST	Digit Symbol substitution Test
OLS	Ordinary Least Squares
PCA	Principal Component Analysis
SVC	Support Vector Classifier
SMOTE	Synthetic Minority Oversampling Technique
ROC AUC	Area Under the Curve of Receiver Operating Characteristic
PPV	Positive Predictive Value
PS AUC	AUC of PPV-Sensitivity curve, a.k.a. precision-recall curve

Chapter 1

Introduction

Alzheimer's disease (AD) is a progressive neurodegenerative disorder that accounts for 60-80% of dementia cases (Garre-Olmo, 2018), known for its insidious and eventually severe cognitive, behavioural and neuropsychiatric symptoms (López and DeKosky, 2008). The management of these symptoms places enormous financial and emotional burden on patients, healthcare systems, healthcare workers and informal carers (Castro et al., 2010; Mahoney et al., 2005; Lowin et al., 2001). Moreover, the prevalence of AD is estimated to double or triple by 2050 from 2019 figures (Nichols et al., 2022), leading the World Health Organisation (WHO) to describe it as a growing epidemic and global health priority (WHO, 2023; Winblad et al., 2016). Only recently have potential disease-modifying treatments resulted from experimental interventions (Sims et al., 2023; Vitek et al., 2023), albeit with important caveats around efficacy and generalisability (Manly and Deters, 2023). One likely explanation for the many failed pharmacological trials to date is that treatments are started too late in the progression of the disease (Mehta et al., 2017; Laske, 2014). Therapies may be more effective if targeted in earlier stages of the disease before significant brain damage has occurred (Guest et al., 2020). Therefore, a major research goal is to investigate methods of earlier AD diagnosis (Whitehouse and George, 2016).

Impairments in spatial navigation and memory are among the earliest detectable cognitive symptoms of AD, caused by pathophysiological damage in the medial temporal lobe (MTL), including entorhinal cortices and hippocampi (Serino et al., 2014). Improving the detection of spatial cognition impairments

may improve the detection of this early brain damage, and hence AD itself. An increasingly popular way of testing spatial cognition is through immersive virtual reality (iVR), technology that presents a simulated three-dimensional environment, giving the impression that the user has stepped into a computer-generated world by allowing six mechanical degrees of freedom (see Figure 1.1). This technology allows naturalistic user interaction with realistic inputs for spatial processing (optic flow, proprioceptive, and vestibular feedback) while allowing the experimenter complete control over environmental features and automated collection of behavioural data. These advantages have already allowed an existing iVR spatial task to detect participants with early cognitive markers of AD (Howett et al., 2019; Castegnaro et al., 2023).

There is scope for development of further iVR spatial AD tests based on current understanding of space and memory processing. This thesis will combine three spatial test components for AD diagnosis into a single iVR paradigm. First, a viewpoint-shifting object-location memory task was translated to an iVR platform for potential AD diagnostic assessment. Testing object-location recognition from a shifted viewpoint has been used to study spatial memory since Piaget and Inhelder first documented young children's difficulty transforming spatial information (Piaget and Inhelder, 1956). Since then, a classic *spatial updating* paradigm involving either table or viewer rotation has formed the basis of numerous experiments to understand object-location memory from a shifted viewpoint (Simons and Wang, 1998, Wang and Simons, 1999; Burgess et al., 2004; Negen et al., 2018; Hilton et al., 2020; Segen et al., 2021; Heywood-Everett et al., 2022). However, this viewpoint-shifting task has not been utilised in AD studies (although see Chan et al., 2016; Chen et al., 2019), despite its utility in testing spatial memory. This thesis will primarily describe the development of an adapted iVR spatial updating task for testing early signs of AD.

An unexplored domain in spatial cognition in AD is the accuracy of subjective judgements of spatial memory performance. A body of evidence sug-

gests a dysfunction in the ability of AD patients to accurately monitor their memory in other domains (Souchay et al., 2002; Dodson et al., 2011; Yu et al., 2020; Li, Sun, et al., 2022; Li, Pan, et al., 2022). However, testing of spatial memory-monitoring has not been translated into AD diagnostics, let alone from a viewpoint-shifting object-location task. A simple confidence rating was introduced to the spatial updating iVR task to assess whether spatial memory monitoring could add value to early AD diagnosis.

Finally, measurement of spatial memory in AD can be enhanced through the use of eye-tracking (Bueno et al., 2019). In experimental literature, tracking eye movements has proven to be a powerful way of exploring encoding and retrieval processes during memory tasks, providing insights into how ocular and attentional mechanisms influence or reflect mnemonic processing (see Ryan et al., 2020 for a review). For example, a small number of studies have used eye-tracking to investigate memory encoding behaviour in healthy younger and older adults in spatial updating tasks (Segen et al., 2021, Hilton et al., 2020). Additionally, impressive diagnostic accuracy for AD has been achieved with memory paradigms making use of concurrent eye-tracking (Parra et al., 2022; Zola et al., 2013). However, the combination of spatial updating and eye-tracking has not previously been used to study memory retrieval processes nor test for discrimination of AD.

Accordingly, this thesis will present a new iVR spatial updating paradigm of a viewpoint-shifting task with additional assessment of spatial memory monitoring and concomitant eye-tracking for potential contribution to future diagnosis of AD. The remainder of this chapter will review current understanding of relevant literature for this undertaking and present a more detailed rationale. The first section describes the current state of clinical AD diagnosis and its prodromal stages to set in context the task of identifying new cognitive diagnostic tools for the disease. Following this, a review of spatial cognition, memory-monitoring, and eye-tracking literature is presented in relation to AD. The spatial updating

paradigm will be introduced, with a rationale for hosting such a task in iVR. Finally, aims and predictions for outcomes of healthy ageing and early AD groups will be described.

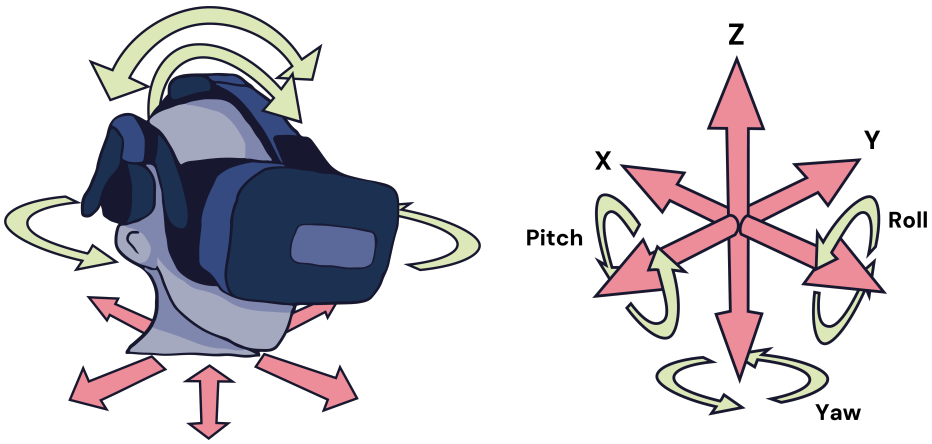


Figure 1.1. Illustration of the six mechanical degrees of freedom provided by immersive virtual reality. The technology usually involves wearing a head-mounted display, allowing translation along and rotation around three dimensions in space. Pink straight arrows represent translational motion through three-dimensional space, green curved arrows represent rotational movement. Note that some forms of iVR only provide three degrees of freedom by allowing head rotations without translational movement. Headset modelled on the HTC Vive Pro Eye used in this thesis.

1.1 Alzheimer's disease and its clinical diagnosis

AD is characterised by a progression of cognitive and behavioural symptoms that usually develop later in life. The majority of cases are sporadic with unknown aetiology, although increased likelihood of diagnosis comes with greater age, lower education and presence of genetic risk factors (Povova et al., 2012; Armstrong, 2019). There are also familial, early-onset types of AD, which often have clear genetic markers that can diagnose the presence of the disease much earlier than sporadic forms. For the purposes of this thesis, the goal of earlier AD diagnosis refers to the sporadic form, which does not have a reliable early marker of the disease. Nevertheless, both sporadic and familial AD have very similar early cognitive symptoms, leading to behavioural problems such as an inability to navigate a well-trodden environment (Pai and Lee, 2016; Tu and Pai, 2006; Passini et al., 1995), or forgetting familiar details of people, places, and events (Bature et al., 2017; Almkvist et al., 1998). With progression of the disease, neurocognitive impairment becomes global, underpinned by cerebral atrophy of up to 40-70% brain volume loss (Schott et al., 2003). Consequently, in moderate to severe stages of AD, patients cannot easily perform activities of daily living without caregiver support (Livingston et al., 2020)

In clinical settings, AD is most commonly diagnosed by testing episodic memory on traditional neuropsychological tests (NICE, 2018) such as the Mini-Mental State Examination (Folstein et al., 1975) or the Montreal Cognitive Assessment (Nasreddine et al., 2005). Indeed, memory dysfunction was the first cognitive symptom to be documented by Alois Alzheimer in his original description of the disease (Alzheimer, 1906), and has subsequently become the most widely-recognised phenotype of AD (Cahill et al., 2015). Numerous studies have described early impairment in recalling verbal or visuospatial information (Greenaway et al., 2006; Almkvist, 1996; Grossi et al., 1993; Martin et al., 1985), which is used to differentiate between AD and other causes of dementia (Dierckx et al., 2007; Estévez-González et al., 2003; Swainson et al., 2001).

Reduced cognitive function is a common characteristic of ageing without dementia (Deary et al., 2009), and detection of neurodegenerative disease is usually measured in comparison to this process. When a patient shows evidence of cognitive impairment which does not preclude activities of daily living, they are diagnosed with a pre-dementia state known as mild cognitive impairment (MCI). This is a heterogeneous category of diagnoses identified through cognitive tests, whereby between 40-60% of patients will progress to a diagnosis of dementia within a few years (Geslani et al., 2005; Petersen, Stevens, et al., 2001; Petersen, Doody, et al., 2001; Petersen et al., 1999). If the neuropsychological profile of these participants reflects a specific impairment in memory function, they will often be diagnosed with so-called amnestic MCI (aMCI). This group has a higher conversion rate to AD than other MCI subtypes (Fischer et al., 2007), and is considered in both clinical and research settings to indicate prodromal AD.

By the time episodic memory impairment is detectable by standardised neuropsychological tests, AD pathology is likely to be wide-spread throughout the brains of patients (Petersen, 2009). The accompanying neurodegenerative damage is irreversible, and may be too late to halt, let alone reverse. This is the reasoning behind recent efforts to diagnose the disease earlier than currently-detectable episodic memory loss, which some believe could significantly contribute to effective treatments (Guest et al., 2020; Mehta et al., 2017; Laske, 2014).

There is ample evidence that the underlying neuropathology of AD predates the earliest cognitive impairments by many years (Jack et al., 2019). Early biological markers of AD brain changes are hence common targets for new diagnostic tools, and indeed there has been a recent push towards biological characterisation and diagnosis of the disease (Jack et al., 2018; Selkoe, 2011). Methods of clinically-approved testing for Alzheimer's biomarkers are currently aimed at quantifying levels of two key AD-related proteins in the cerebrospinal

fluid: amyloid-β_{42}-peptide (Aβ) and phosphorylated tau (p-Tau). These proteins indicate the presence of neuritic plaques and neurofibrillary tangles, respectively, that have both been connected to synaptic loss and nerve cell death in the disease (Hardy and Selkoe, 2002; Thal et al., 2002; Price and Morris, 1999; Braak and Braak, 1991). These biological hallmarks of the disease have contributed to the establishment of the so-called Amyloid/Tau/Neurodegeneration (ATN) Framework in Alzheimer's disease diagnosis (Jack et al., 2018). This conceptualisation of the disease aims to provide a common definition of its biological underpinnings for research purposes, highlighting the interaction between Aβ, Tau and neuronal damage.

Currently, presence of Alzheimer's proteinopathies can be clinically investigated in the United Kingdom through one of two means: (1) a lumbar puncture and subsequent test of cerebrospinal fluid (CSF) for Aβ and p-Tau, and (2) Positron Emission Tomography (PET) of these proteins (NICE, 2018). Generally, thresholds for protein quantities are used to inform differential diagnosis, such that a patient with MCI and presence of Aβ and/or p-Tau may be categorised as biomarker positive (MCI+), and therefore in the prodromal stages of AD. By contrast, biomarker negative MCI patients (MCI-) are more likely to have suffered cognitive impairment for an alternative reason, such as depression, anxiety, another type of dementia, cerebrovascular disease, or normal age-related decline (Wang et al., 2021; Wisse et al., 2015). Clinical examination of AD biomarkers is therefore a highly valuable procedure, recommended as a supportive component to early AD diagnosis (NICE, 2018; Wolfsgruber et al., 2017).

Unfortunately, the availability of clinical AD biomarker testing is mired by practical barriers. Lumbar punctures to acquire CSF samples are invasive procedures requiring trained specialists, with variability between laboratories leading to testing biases (Mattsson et al., 2010). Moreover, if a patient cannot undergo a lumbar puncture for any reason, the alternative PET scan requires ingestion of a radioactive tracer, which reduces patient acceptability despite its

relative safety (May et al., 2020). Furthermore, PET is a highly expensive scanning technique, with low availability due to its rarity (McMahon et al., 2003; Wittenberg et al., 2019). These limitations likely contribute to low diagnosis rates, which are estimated at 20-50% in the UK and just 10% in developing nations (Patterson, 2018).

Several alternative means of testing for AD biomarkers have been proposed, but are currently only available in research settings. For instance, blood-based biomarker testing has shown high predictive validity for CSF protein levels (Gao et al., 2023; Varesi et al., 2022) via a relatively cheap, available and less invasive procedure. Even less invasive biomarker testing includes saliva samples (Paraskevaidi et al., 2020) and breath-based tests (Emam et al., 2020). With further investigation, standardisation, and regulation, these forms of biomarker testing could reduce the cost and improve the availability of Aβ and p-Tau detection in clinical practice.

Despite developments in biomarker testing, research into new early *cognitive* markers of AD is still important for a number of reasons. For one, neuropsychological impairment can directly relate to the clinical symptoms that impact on people's ability to function in daily life, which biological markers aim to predict (Livingston et al., 2020). For example, an impairment in spatial processing may lead to people with AD getting lost in a familiar environment (Coughlan et al., 2018). Cognitive markers of AD also play an important role in the development of new treatments as outcome measures for intervention effectiveness (Andrews et al., 2019; Mehta et al., 2017).

Moreover, protein and genetic biomarkers do not currently have sufficient predictive validity for diagnosis because many people develop significant proteinopathy and neurodegeneration without cognitive symptoms (Dubois et al., 2021; Livingston et al., 2020; Jack et al., 2019). For example, a 65 year-old with positive amyloid biomarkers but intact cognition has just a 2.3-2.5% risk of developing AD in the following 10 years (Brookmeyer and Abdalla, 2018). Clini-

cal diagnosis of AD still relies on neuropsychological assessment, which partly explains why the aforementioned ATN network is not recommended for medical diagnosis, as a biological definition of AD separated from cognitive symptoms is not yet accurate enough for clinical identification (Jack et al., 2018). Accordingly, developing more sensitive tools for cognitive impairment is still important for earlier identification of the disease.

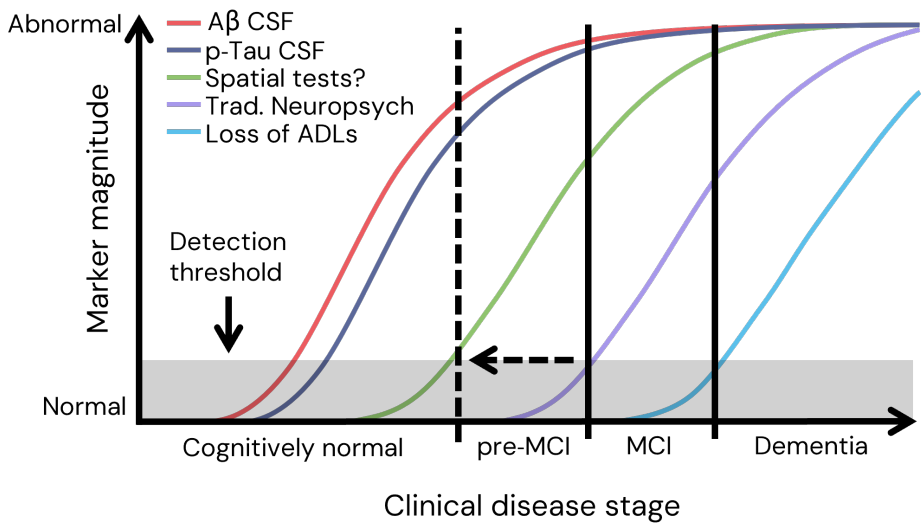


Figure 1.2. Model of biological, cognitive and functional markers through different stages of Alzheimer's disease progression. The green line represents a hypothesised potential early dysfunction in spatial cognition, which may allow earlier diagnosis of MCI (dotted black line and arrow) compared to traditional neuropsychological tests (Trad. Neuropsych). CSF biomarkers would pre-date this but do not have sufficient predictive validity alone (see main text). Aβ: amyloid-beta; p-Tau: phosphorylated tau; CSF: cerebrospinal fluid; ADLs: activities of daily living. Adapted from Jack et al., 2010 and Jack, 2022.

1.2 Spatial cognition as a cognitive marker of AD

Novel cognitive markers of AD will likely be successful if they place high neurocognitive demand on brain regions with early neuronal damage in AD. Functions of the medial temporal lobe (MTL) are appropriate targets for this (Mori et al., 1997; Cavedo et al., 2014; Jack et al., 1998): in early stages of the disease, amyloid plaques and neurofibrillary tangles accumulate in the hippocampal formation and entorhinal cortices (Price and Morris, 1999; Thal et al., 2002; Braak and Braak, 1991). These depositions have been specifically linked to neuronal damage (Bloom, 2014; Cárdenas et al., 2012) with volume reductions detectable in MTL areas by structural MRI scans of AD patients' brains (Schott et al., 2003). Furthermore, recent cases of resistance and resilience to familial AD have specifically implicated intact entorhinal cortices in significantly delayed memory loss (Lopera et al., 2023; Arboleda-Velasquez et al., 2019). These results suggest that identification of early AD may be achieved by testing the cognitive functions underpinned by MTL areas.

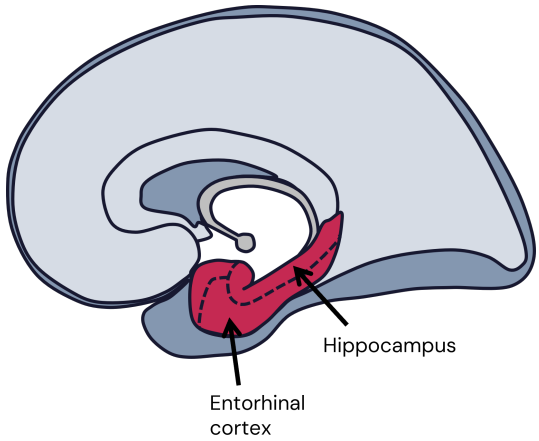


Figure 1.3. Basic illustration of medial saggital view of the human brain, with the medial temporal lobe highlighted in red. The hippocampus and entorhinal cortex are labelled, with dotted lines representing potential borders; labels of other areas of the MTL are omitted, as are details of the wider brain. Relative sizes not precisely to scale.

1.2.1 Spatial cognition and the medial temporal lobe

Areas of the MTL have been heavily implicated in spatial memory and navigation (Bird and Burgess, 2008; Burgess et al., 2002). An important area of research on the neural underpinnings of these functions has been the identification of several specialised groups of cells with spatially-modulated firing patterns, such as place cells and grid cells (Hartley et al., 2014). Place cells, found mainly in the hippocampus, fire when an animal is in a specific location in an environment, independent from the direction it is facing (O'Keefe, 1976). Collectively, these cells' corresponding place fields provide a cognitive representation of the animal's environment, or 'cognitive map' (O'Keefe & Nadel, 1978). Grid cells, discovered mainly in the entorhinal cortex (Hafting et al., 2005) fire when an animal is located at a series of different locations in an environment, organised in a grid-like arrangement (Grieves, Roddy & Jeffery, 2017). These and other specialised cells have been shown to be critical for navigation and representation of the environment in non-human animal studies (Hartley et al., 2014).

Evidence supports the existence of spatially-specialised cells in humans as well (Epstein et al., 2017). For example, functional magnetic resonance imaging (fMRI) studies have demonstrated a six-fold grid cell-like signal during navigational tasks that is consistent with the arrangement of grid cell firing discovered in non-human animals (Kim and Maguire, 2019; Doeller et al., 2010). Moreover, in-vivo neurophysiology studies have directly recorded firing patterns from human MTLs that are consistent with place cells (Ekstrom et al., 2003; Jacobs et al., 2010; Miller et al., 2013), and grid cells (Jacobs et al., 2013; Maidenbaum et al., 2018). Case studies of people with MTL lesions have also shown specific impairments in spatial functioning, including navigation and topographical difficulties (Bird et al., 2007; Cipolotti et al., 2006; Spiers et al., 2001), impaired recollection of scenes (Bird et al., 2008), and reduced memory for object locations from a shifted viewpoint (King et al., 2002).

Although the MTL's involvement in spatial processing is well established,

spatial cognition is not exclusive to this brain region. A distinction is often made between two 'frames of reference' used to represent spatial information that are tied to separate, but highly connected, regions of the brain. *Egocentric* representations reflect subject-to-object relationships, important for movement through an environment because navigation requires representation of the body in relation to environmental cues (Colombo et al., 2017). Evidence suggests this type of processing is underpinned by the caudate nucleus and areas of the medial and posterior parietal lobe (Cook and Kesner, 1988). *Allocentric* processing, in contrast, is object-to-object, world-centred representation, independent from the subject's point of view and centred on environmental landmarks (Ekstrom et al., 2014; Klatzky, 1998). This relies on processing in the hippocampus, particularly hippocampal place cells (Ekstrom et al., 2003), which are supported by head-direction cells for orientational information (Taube, 2007; Taube, 1998) and entorhinal grid cells for path integration (Hafting et al., 2005). Successful navigation and memory of spatial environments requires representation and retrieval of both egocentric and allocentric frames (Burgess, 2006; Burgess, 2008). However, the allocentric frame has been more closely connected to the MTL and is therefore a good target for testing damage to this brain area (King et al., 2002).

Several previous studies have demonstrated impairments in allocentric spatial memory in early AD. One test required participants to identify an arrangement of four mountains from a shifted viewpoint (see Figure 1.4; Hartley et al., 2007). This 4 Mountains Test (4MT) has demonstrated differential discrimination of MCI participants with and without evidence for Alzheimer's biomarkers (Chan et al., 2016). Another study suggested a specific impairment in allocentric distance judgement may be present in AD during an object-location memory task (Ruggiero et al., 2020).

Studies focusing on spatial navigation abilities, rather than memory of external object positions, have also found results consistent with an allocentric im-

pairment in AD. For example, human real-world and virtual maze studies have found that aMCI participants were significantly impaired on allocentric subtests (Hort et al., 2007; Laczó et al., 2014; Lee et al., 2014). Likewise, virtual (Howett et al., 2019) and blindfolding (Mokrisova et al., 2016) path integration tests have found impaired navigation performance in AD populations. It should be noted that several spatial navigation studies in AD have also found egocentric impairments (Laczó et al., 2014; Weniger et al., 2011; Serino et al., 2015). This may be explained by a specific impairment in the transition between the two reference frames; two studies have found AD-specific impairments in egocentric-allocentric switching and syncing (Serino et al., 2015; Ruggiero et al., 2018), leading some authors to suggest an impairment in translating spatial information from an allocentric to egocentric representation (Serino and Riva, 2013).

Although the specific mechanism of impairment is still under investigation, allocentric tasks have now consistently presented difficulties for participants with AD, supporting this type of spatial memory for diagnostic testing of the disease. The spatial updating task developed in this thesis was based on this understanding, with significant allocentric components included.

1.2.2 Self-motion

Another key component of spatial processing is the updating of spatial information as an animal moves through the environment (Taube, 2007; Kropff et al., 2015; Shine et al., 2016). Self-motion cues such as optic flow, vestibular feedback, and motor efference copy are used to track the position of the animal (Ellmore and McNaughton, 2004; Angelaki, 2014; Sherrill et al., 2015) and external surroundings (Chan et al., 2012). Importantly, studies implicate the entorhinal cortex in self-motion calculations of an animal's position in space (Campbell and Giocomo, 2018; Mallory et al., 2021). A spatial test may therefore be sensitive to AD damage if entorhinal self-motion processing is required for task performance (as in Howett et al., 2019).

Figure 1.4. (Left) diagram illustrating the difference between egocentric and allocentric representations; (Middle) an example stimulus from the 4 Mountains Test. Participants must memorise an arrangement of four mountains like this one. After a few seconds, the scene disappears and a forced alternative choice of four landscapes is shown, with only one matching the first mountains but from a shifted viewpoint. Reproduced from Hartley et al., 2007; (Right) Diagram of a trial in the Path Integration Test (PIT; Howett et al., 2019), which requires participants to walk two sides of a triangle, and try to complete it by returning to the starting position. MCI+ participants have greater absolute distance and angular error than MCI- participants. Figure reproduced from Castegnaro et al., 2023.

The importance of self-motion in healthy allocentric memory has been demonstrated in experiments requiring participants to remember object locations from a shifted viewpoint. Several studies have shown that self-motion to a new viewpoint led to better memory performance than shifting the viewpoint without self-motion (Chance et al., 1998; Simons and Wang, 1998; Burgess et al., 2004; Holmes et al., 2018; Tascón et al., 2018). In this *spatial updating* paradigm, encoded object locations remain consistent with self-motion if a participant walks to a new viewpoint while the objects remain stable (see Figure 1.5). Alternatively, object locations are inconsistent with self-motion if the participant's viewpoint remains the same and the configuration of objects rotates instead. Simons and Wang demonstrated that memory performance from a shifted viewpoint was greater with self-motion consistency than without (Simons and Wang, 1998). Burgess, Spiers and Paleologou extended this by demonstrating that the performance-enhancing effect of self-motion consistency was dissociable from—but additional to—that of allocentric consistency: by varying consistency of object locations with the position of an external cue (Figure 1.6),

they showed that participants detected a moved object better when the configuration of the objects was consistent with either self-motion or external cues, and a greater boost when both were consistent (Burgess et al., 2004). This suggests that self-motion processing independently contributes to object-location representations, and an impairment in this functionality may contribute to difficulties in spatial memory.

The benefits of self-motion to object-location memory have not been tested before in early AD, but research suggests an impairment in self-motion processing may be present in the disease. For example, self-motion is a key component of path integration (Evans et al., 2016; Campbell and Giocomo, 2018), which has been specifically impaired in AD participants when tested using iVR (Howett et al., 2019) and real-world blindfolding (Mokrisova et al., 2016) paradigms as previously mentioned. Indeed, path integration impairment has been linked to entorhinal cortex damage in mouse models of AD (Ying et al., 2023). However, it is worth noting that one study also found intact path integration performance from participants with hippocampal or entorhinal lesions (Shrager et al., 2008).

In addition to path integration, research suggests an impairment in optic flow functions during self-motion perception in AD participants (Wang, Guo, et al., 2017), and an increased likelihood of vestibular dysfunction (Agrawal et al., 2020). Taken together, previous results suggest that the major components contributing to self-motion updating of spatial representations may have been impaired in AD studies. Therefore, the benefit of self-motion to object-location memory may also be impaired in AD patients.

Accordingly, the spatial updating task presented in this thesis will include variation of self-motion to test for AD-related impairments in this functionality, and potentially provide better diagnostic accuracy for early AD than if focusing on allocentric memory alone. The details for how self-motion was manipulated are provided in Section 1.5 of this Chapter. Before then, two further extensions to this paradigm are introduced and explained: memory monitoring and eye-

tracking. The following two sections will outline the main reasons for including these adaptations.

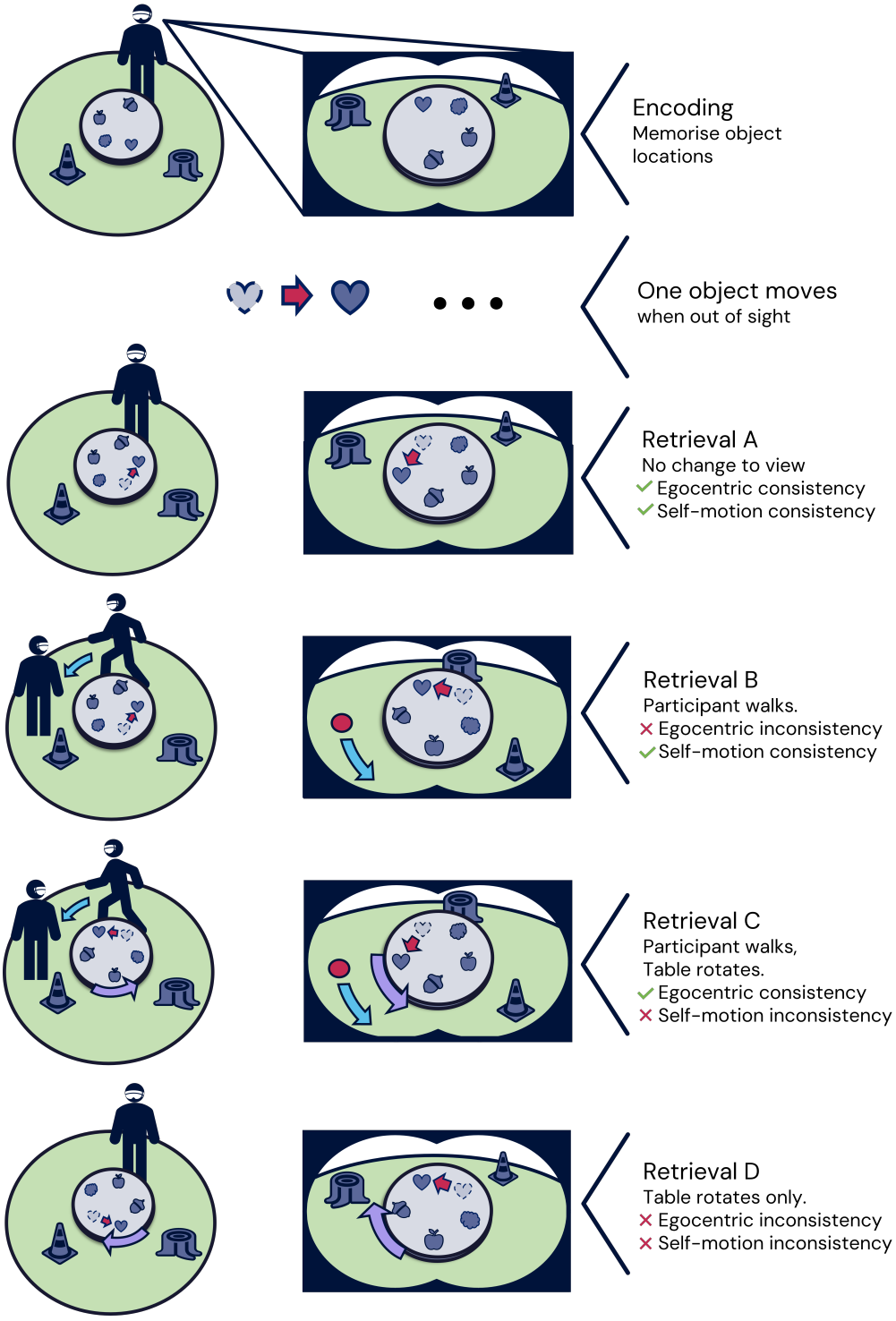


Figure 1.5. Cartoon demonstrating the manipulation of egocentric and self-motion consistency in the spatial updating paradigm. Left-hand figures show a third-person schematic, right-hand figures show the first-person view. The participant views the object from the first viewpoint (encoding, top); during a period of occlusion where the participant cannot see the table, one object moves; four different retrieval conditions are created by rotating either the participant or table around the centre of the table.

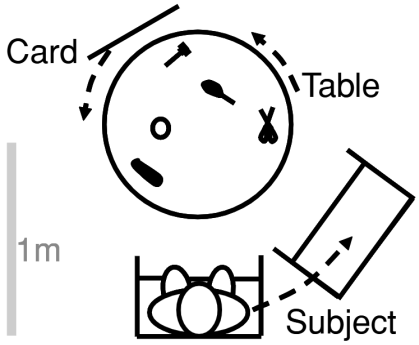


Figure 1.6. Diagram of the spatial updating paradigm including variation of consistency with an external cue. The same process as Figure 1.5 applies, but with another factor: an external cue was manipulated to be consistent or inconsistent with the encoded positions of the objects. Reproduced from Burgess, 2008.

1.3 Memory monitoring

Another area of cognition with potential for detection of AD is the assessment of subjective memory-monitoring abilities. Memory-monitoring, or 'metamemory', is defined broadly as people's knowledge of their own memory (Chua et al., 2014). This can be tested more specifically via memory monitoring accuracy, which refers to people's ability to correctly judge the accuracy of their objective memory performance (Lam et al., 2012). There is evidence for an impairment in episodic memory monitoring in participants with AD (Dodson et al., 2011; Souchay et al., 2002): patients were poor at judging the accuracy of their recognition memory, even when their performance was the same as healthy controls. In later stages of the disease, this dysfunction may contribute to a more global unawareness of cognitive dysfunction known as anosognosia (Galeone et al., 2011). However, studies have also found metamemory impairments in the earliest stages of cognitive impairment by documenting self-monitoring inaccuracies in participants with subjective cognitive decline (Li, Sun, et al., 2022; Yu et al., 2020; Li, Pan, et al., 2022)—a preclinical diagnosis defined by subjective experience of cognitive impairment without a measurable reduction in memory perfor-

mance (Jessen et al., 2020). These results suggest that impairments in memory monitoring may be detectable earlier than objective memory dysfunction.

However, other results suggest that memory monitoring in early AD remains intact. For example, studies have shown no difference in confidence ratings of answers to general knowledge questions by AD participants compared to healthy controls (Bäckman and Lipinska, 1993; Lipinska and Bäckman, 1996; Cosentino et al., 2007), indicating an intact ability to monitor semantic memory performance. Similar findings have been found in prodromal stages of the disease (Chi et al., 2022), suggesting that semantic—but not episodic—memory monitoring is spared early in the disease. However, intact *episodic* memory monitoring has also been found in AD patients compared to healthy controls (Gallo et al., 2012), despite findings suggesting the opposite, as described in the previous paragraph. Furthermore, some studies have found reduced episodic memory monitoring abilities in healthy ageing in the absence of neurodegenerative disease (Comblain et al., 2004; Dodson et al., 2007 ; Johnson et al., 2015; Angel et al., 2022). The reason for these discrepancies may lie in the results of one study, which found that impaired metamemory was specifically related to nonverbal episodic memory (Cosentino et al., 2007). These results suggest that the type or sub-type of memory may be important for detecting impairments in memory monitoring abilities in AD, dissociable from healthy age-related changes.

To the best of my knowledge, no previous studies have assessed spatial memory as a sub-type of memory monitoring in AD or ageing. However, evidence suggests that variability in memory monitoring in AD patients may be related to specifics of neuroanatomical decline (Cosentino et al., 2007). Furthermore, memory monitoring impairment in prodromal stages of AD has been linked to cortical thinning and neural activity alterations in the MTL (Li, Sun, et al., 2022). Considering that spatial memory impairment related to MTL changes has been a consistent finding of early AD (see sections 1.2.1 and 1.2.2), spa-

tial memory monitoring is worth investigating in AD to test whether this domain of metamemory is particularly useful in detecting the disease. Accordingly, a memory monitoring component was included in the iVR spatial task as a potential contributor to AD diagnosis.

1.4 Eye movements as a marker of AD

Eye tracking is a promising, powerful measure of cognitive functioning for spatial testing in AD, in part due to the connection between eye movements and the MTL. Indeed, there is evidence for analogous specialised cells for eye movement 'navigation' as for whole-body spatial navigation (Nau, Julian, and Doeller, 2018). For example, entorhinal cortex visual grid signals have been shown to be anchored to visual space, similar to how grid cells represent navigable space (Killian et al., 2012). Furthermore, Nau and colleagues provided MRI evidence of a hexadirectional visual grid signal in the entorhinal cortex during a controlled gaze direction task, similar to the pattern of firing during spatial navigation (Nau, Navarro Schröder, et al., 2018). Similarly, visual border cells have been found to fire near the border of visual space, analogous to boundary vector cells in spatial navigation (Killian et al., 2012).

In more naturalistic paradigms, a well-documented 'preferential viewing' effect has been related to MTL functioning. This phenomenon refers to the increased likelihood of participants viewing novel or changed stimuli (Manns et al., 2000; Ryan et al., 2020), essentially demonstrating a memory effect on eye movements. For example, visual fixations were greater for objects that had been moved within a previously viewed scene (Yeung et al., 2019). This was associated with increased entorhinal cortex volume (Yeung et al., 2020). Moreover, humans and primates with lesions to the MTL exhibited a reduced preferential viewing effect (Zola et al., 2000; Ploner et al., 2000; Hannula et al., 2007; Lucas et al., 2018), suggesting that memory processing in the MTL influences eye movements.

Eye movements have also been more explicitly linked to memory performance. For example, when participants were free to explore a visual scene, fixation patterns predicted subsequent memory performance (Fehlmann et al., 2020; Olsen et al., 2016; Loftus, 1972) as well as activity in the MTL (Liu et al., 2017; Fehlmann et al., 2020). Indeed, several studies have found that recognition of previously viewed visual stimuli was improved when participants looked at the same features at encoding and retrieval (Mäntylä and Holm, 2006; Holm and Mäntylä, 2007; Foulsham and Kingstone, 2013), even showing recapitulation of fixation sequences made during first viewing (Wynn et al., 2018; Wynn et al., 2016; see Wynn et al., 2019 for a review). This *gaze reinstatement* effect has been connected to areas of the MTL including the hippocampus and parahippocampal gyrus in fMRI studies (Wynn et al., 2022; Ryals et al., 2015), supporting the theory that eye movements themselves are embedded into memory representations (Bicanski and Burgess, 2020; Bicanski and Burgess, 2019; Wynn et al., 2019). Accordingly, dysfunction in the MTL may disrupt gaze reinstatement, although this explicit connection has not yet been studied.

The above pattern of results suggests that eye tracking would be sensitive to Alzheimer's-related damage, and indeed eye movement changes have been observed in patients with MCI and AD. For example, key differences in saccades (fast movements between periods of fixated gaze) have been observed in AD through prosaccade and antisaccade paradigms (see Opwonya et al., 2022 for a review), which require the participant to intentionally saccade towards or away from a target stimulus, respectively. AD patients performing these tasks have exhibited reductions in saccadic velocity, accuracy and initiation speed compared to healthy control volunteers (Molitor et al., 2015).

Eye movement differences have also been found during memory-related paradigms in patients with MCI, AD and MTL damage. For instance, research has supported a reduced preferential viewing effect in AD (Crutcher et al., 2009), which elsewhere predicted conversion from MCI to AD (Zola et al., 2013). In-

deed, these studies showed that AD patients would fixate less on incongruous or changed areas of visual scenes, indicating a reduced memory effect on eye movements.

Preferential viewing effects are necessarily only observable during memory retrieval (i.e. after the first viewing of the stimulus). Additionally, differences in how early AD participants *encode* information during memorisation of visual stimuli have been found. For example, one study found that participants at genetic risk of familial AD spent less time fixating the stimuli during learning, interpreted as contributing to ineffective encoding strategies (Pavisic et al., 2021). This same study found that symptomatic AD participants showed greater inequalities in the distribution of their fixation patterns across visuospatial stimuli. This finding is somewhat in contrast to studies that have found a more random—and therefore equal—distribution of fixation, as measured by transition entropy, in participants with MCI (Coco et al., 2021) and MTL damage (Lucas et al., 2018). Still, these findings indicate that an encoding dysfunction related to MTL damage is detectable via eye-tracking. This is one example of the utility of eye-tracking over memory performance alone, which cannot easily dissociate between encoding and retrieval impairments.

Utilisation of eye-tracking in spatial memory paradigms has been implemented, but only rarely, despite the importance of vision for spatial processing (Lester et al., 2017; Burkhardt et al., 2023). Indeed, eye-tracking can reveal mnemonic differences where declarative memory cannot. For example, one study used eye tracking to examine gaze patterns during object-location memory and found that viewing patterns during spatial location encoding in healthy younger and older adults were different despite equal performance on the memory task (Segen et al., 2021). Specifically, older adults would attend to more information during encoding than younger adults. A similar study found that older adults had different patterns of fixation sequences when memorising object locations (Hilton et al., 2020). These studies demonstrate the power of

eye-tracking in spatial memory tasks, which has rarely been included elsewhere (see also Coco et al., 2023). However, these authors only examined encoding patterns i.e. eye movements during the learning phase of the memory task. There is a gap in the literature for examining eye movements during spatial memory *retrieval*, especially after a shifted viewpoint.

To summarise, tracking eye movements during spatial memory could be a promising approach for testing AD. Indeed, several studies have shown diagnostic discrimination of the disease using eye movements alone (Zola et al., 2013; Pavisic et al., 2021; Parra et al., 2022). Some of these have involved spatial elements by virtue of including visuospatial stimuli (e.g. Pavisic et al., 2021). However, the combination of eye tracking for studying spatial memory under different spatial reference frames in AD is scarce. Previous relevant studies have only presented visuospatial tasks from an egocentric reference frame through the use of two-dimensional desktop tasks (Coco et al., 2023), or avoided gaze analysis during retrieval phases (Segen et al., 2021; Hilton et al., 2020). Using eye-tracking to examine memory retrieval patterns during allocentric memory has not been done before, but could be a promising way of measuring spatial impairment in AD.

1.5 A spatial updating paradigm in immersive virtual reality

The *spatial updating* task used by several others to investigate spatial memory processing in healthy adults (see Section 1.2.2) can be adapted to examine spatial memory performance in early AD. This paradigm allows variation in the use of egocentric and allocentric strategies by shifting the participant's viewpoint of a configuration of objects between encoding and retrieval.

In the task, participants view a table of objects twice, separated by a period of occlusion. On the second viewing, they must detect a spatial change, such

as a moved object. By manipulating the rotation of the participant or the objects around the centre of the table between viewings, the availability of egocentric strategies can be varied. For example, if there is no change between viewings, or if both the participant and the table have rotated by the same amount around the centre of the table, then an egocentric visual 'snapshot' can be used to detect any changes (Simons and Wang, 1998; Burgess et al., 2004). However, if the participant moves to a new viewpoint for the second viewing without the table rotating, they cannot use egocentric strategies but can use object-to-object (allocentric) relative positions to solve the task (see Figures 1.5 and 1.6). Moreover, if the participant walks to the new viewpoint, then the positions of the objects are consistent with their self-motion cues. Indeed, the act of self-motion has been shown to contribute to memory performance in the spatial updating task when a viewpoint is shifted (Simons and Wang, 1998; Burgess et al., 2004). Accordingly, walking to a new viewpoint involves both allocentric and self-motion processing (Burgess et al., 2004) that could be impaired in participants with AD due to MTL damage.

Teleporting and self-motion. The effects of self-motion on memory performance in this task have previously been varied by including table rotations. As described in Section 1.2.2, the effects of self-motion and egocentric views on spatial change detection can be tested by including two further conditions: (1) the table rotates but the subject stays at the same viewpoint, and (2) the participant moves to a new viewpoint and the table also rotates. These two conditions provide both same and shifted viewpoints while disrupting the consistency of object locations with the participant's self-motion (Figure 1.5).

For this thesis, the spatial updating paradigm was adapted to disrupt self-motion in a new way. Immersive virtual reality (iVR) has the advantage of allowing both (a) naturalistic self-motion through the virtual environment, providing proprioceptive, optic flow, and vestibular cues that mimic real-world movement, and (b) unnaturalistic instant transposition of the viewing camera, also known as

'teleportation'. This functionality allows the removal of normal self-motion during allocentric viewpoint shifts by having the participant teleport to the new viewpoint instead of walking. Previous research has shown a reduced performance for healthy adults in a spatial updating task when teleporting to a new viewpoint compared to walking (with no rotation of the objects; Castegnaro, 2021, ch. 4, p. 149). This effect could be due to the disruption of self-motion updating of object locations, which normally provides a boost to spatial memory (see Section 1.2.2), implying that teleportation is a useful tool for investigating movement without self-motion. However, for this to be valid, teleporting *and* rotating the table should *not* be more difficult than walking and rotating the table. This is because both conditions remove the consistency of object positions with self-motion, but maintain the egocentric view (because both conditions involve a shift of view and table rotation by the same angle around the centre of the table, see Figure 1.7).

Eye-tracking. Modern iVR systems now include integrated eye-tracking, allowing measurement of eye movements during the spatial updating task. The paradigm also lends itself well to measurement of gaze location because the main visual stimuli (the objects on the table) can be distinctly separated into areas of interest without ambiguity of the foveated object. Indeed, eye-tracking has been used with a similar paradigm to examine encoding strategies in younger and older adults (Segen et al., 2021; Hilton et al., 2020). However, eye movements were not examined during retrieval in these studies because (a) it was not the focus of the hypotheses, and (b) retrieval viewing varied greatly in length due to differing reaction times, making comparisons of eye movement patterns more difficult. Yet, viewing patterns during memory retrieval may reveal important details of task-relevant eye movements, as evidenced by preferential viewing effects (see Section 1.4). To allow for comparison between retrieval phases, a fixed-length period of forced viewing was introduced before allowing selection of the target object.

Spatial memory monitoring. The final adaptation to the paradigm was the inclusion of a confidence rating scale after deciding which object had moved. This was a cheap and easy way to include a measure of spatial memory monitoring: a first for this task in AD participants.

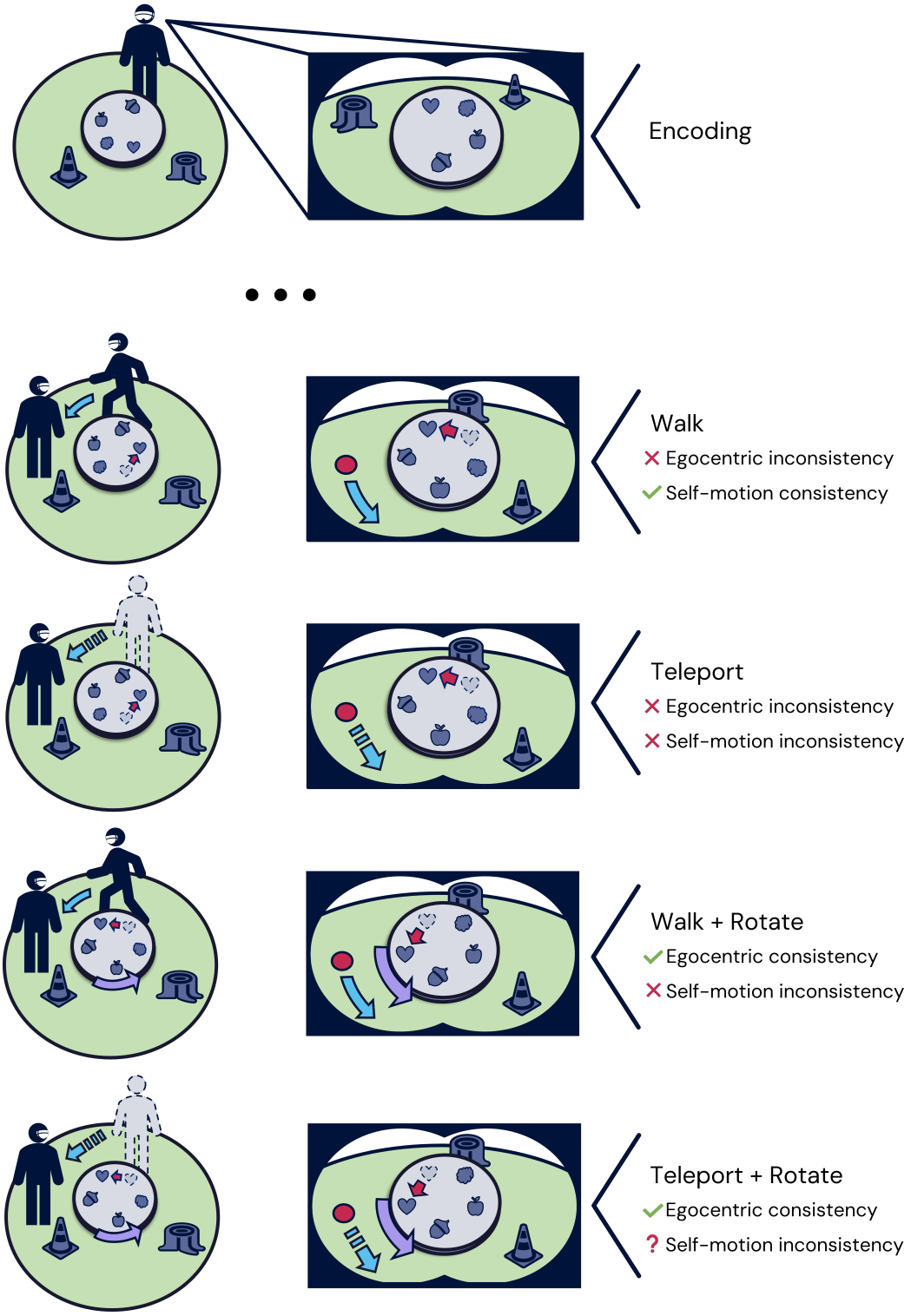


Figure 1.7. Cartoon demonstrating the egocentric and self-motion changes in the spatial updating paradigm if the participant teleports instead of walking. Teleporting disrupts naturalistic self-motion while allowing viewpoint shifts. The implications for this in the Teleport-Rotate condition are not straightforward (hence the question mark), and are discussed further in Section 4.3.2.1 following more detailed hypotheses.

1.6 Aims and predictions

This thesis describes the development of the aforementioned extended viewpoint-shifting spatial updating task in iVR, with additional collection of confidence and eye movement data to measure spatial memory monitoring and memory-related fixation patterns, respectively.

I first aimed to determine the feasibility of this new task due to the inclusion of several new components, including virtual teleportation, eye-tracking, and iVR use by older adults. Indeed, replication of previous spatial memory results in healthy participants was a secondary aim for this adapted task. Although healthy older adults have shown impairments in allocentric navigation tasks (Harris et al., 2012; van der Ham et al., 2020; Zhang et al., 2021), studies of viewpoint-shifting paradigms have *not* found differential effects of shifted perspectives with age (Segen et al., 2021; Hilton et al., 2020). Therefore, older adults were expected to replicate previous patterns of spatial memory results, showing improved performance when object locations were consistent with egocentric views, and allocentric conditions with self-motion via walking. Healthy younger and older adults were included to test the prediction that older adults would perform generally worse overall, regardless of task conditions.

I further aimed to test whether this adapted paradigm would provide useful discrimination of participants with early AD and therefore demonstrate diagnostic utility. This was achieved by recruiting participants with MCI who had received clinical biomarker testing, allowing comparison of those with presence of Alzheimer's biomarkers (MCI+) from those without (MCI-). This approach, while challenging for recruitment, provided a powerful means of dissociating MCI due to AD from other causes in the absence of longitudinal follow up.

Due to the likely impairment of allocentric processing, I expected MCI+ participants to find detection of the moved object more difficult than MCI- or age-matched cognitively normal controls when the task conditions required allocentric processing and self-motion via walking (Retrieval B in Figure 1.5, or

the 'Walk' condition in Figure 1.7). MCI+ participants were predicted to perform worse than control groups in conditions where they walked to a shifted viewpoint, but less so when teleporting to the new viewpoint. Conversely, a specific impairment in allocentric representations and self-motion processing should result in equal performance between MCI+ and MCI- participants when object locations were inconsistent with external cues or self-motion (as in Retrieval C and D in Figure 1.5), and when object movement could be detected with egocentric strategies (Retrieval A and C in Figure 1.5). Note that these condition-specific hypotheses will be revisited and elaborated upon in Chapter 4 prior to the main results of the study.

For eye movements, an iVR system with integrated eye-tracking was used to simultaneously collect eye movement data during the task. I predicted that MCI+ participants would be distinguishable from control participants based on eye movements at encoding and retrieval. Most intuitively, eye movements that have shown a relationship to MTL memory, such as *preferential viewing* of the moved object at retrieval (Yeung et al., 2019) and *gaze reinstatement* compared to the encoding scan-path (Wynn et al., 2019), were predicted to follow the same pattern of results as memory performance.

Conversely, some eye movement measures were predicted to negatively correlate with memory performance and Alzheimer's risk. This may have been the direct inverse of preferential viewing effects, such as fixating on the objects that did not move, or related to the absence of strategic eye movement patterns. Indeed, as mentioned in 1.4, participants with MTL damage had more random or less predictable patterns of fixations on an arrangement of shapes (Lucas et al., 2018; Coco et al., 2021).

For spatial memory monitoring, I predicted that participants with early signs of AD would show less consistency between confidence ratings and memory performance than healthy controls. Indeed, a confidence rating alone does not provide this information, but the relationship between confidence and task per-

formance was compared and contrasted across groups, with the expectation that participants with early signs of AD would show more decoupling between confidence and task performance.

Finally, and perhaps most importantly, I aimed to demonstrate that differentiation of MCI+ participants from controls would be most accurate from task-derived measures compared to traditional neuropsychological tests when training classification models. Further, I predicted that a combination of memory, memory monitoring, and eye-tracking measures would outperform any single measure.

More specific hypotheses and results are described in Chapter 4. Before this, details of the task, its technical development and data pipeline are detailed in Chapter 2. The feasibility of the task for testing younger and older adults is then described in Chapter 3. Following this, the results of the above predictions are presented in Chapters 4-6, which detail the findings from a series of analyses comparing groups on eye movement and behavioural measures.

Chapter 2

Materials and Methods

This chapter contains details of the development of the study paradigm. The virtual reality implementation of the task is described including the processing of eye data. The details in this chapter hold across all successive chapters.

2.1 Materials and setup

The task was developed in the Unity3D game engine (version 2018.3.9f1) and run using the HTC Vive Pro Eye system (HTC, n.d.), which includes a head-mounted display (HMD) with integrated eye trackers. The two included 'Base Stations' were placed in opposite corners of a space at least 2.5m x 2.5m. The HTC Vive Wireless Adapter and accompanying battery pack were attached to the top of the HMD to allow for free movement during the task. This was important because movement and rotation during the task led to twisting and tangling of the cable if the HMD was tethered to the operating computer.

The manufacturer's reported specification of the eye tracking states a 120 Hz eye sample frequency, a 0.5-1.1° accuracy within the middle 20° field of view and a trackable field of view of 110° (HTC, n.d.). Developers can interact with the eye tracking functionality using the SRanipal Software Development Kit (SDK), which was accessed using the Unity3D game engine. Raw data output from the eye tracking included timestamped gaze origin and direction, pupil position and size, and eye openness. Additionally, gaze origin and direction are used by the SRanipal SDK to calculate a ray vector cast from the origin of each eye. The intersection of this vector with a virtual object can be recorded online and

combined into one world-space gaze point per eye sample. These gaze points formed the basis of fixation and saccade calculations, as described in Section 3.2.5.

The virtual environment (Figure 2.1C) was modelled on a mountainous region with irregular distal cues (mountains) and unique, mid-distance cues in four cardinal directions to aid with orientation. In the centre of the environment was a circular table with a coloured border and an opaque dome on top with an arrow shape to show table direction (and therefore rotation). Two objects, a log and a rock, were placed closer to the table and dome to allow for proximal visual cues.

A boundary warning system was programmed into the virtual environment. If a participant moved within 35 cm of the border of the VR 'play area', then an obvious sign appeared indicating that the participant should move back towards the centre of the arena.

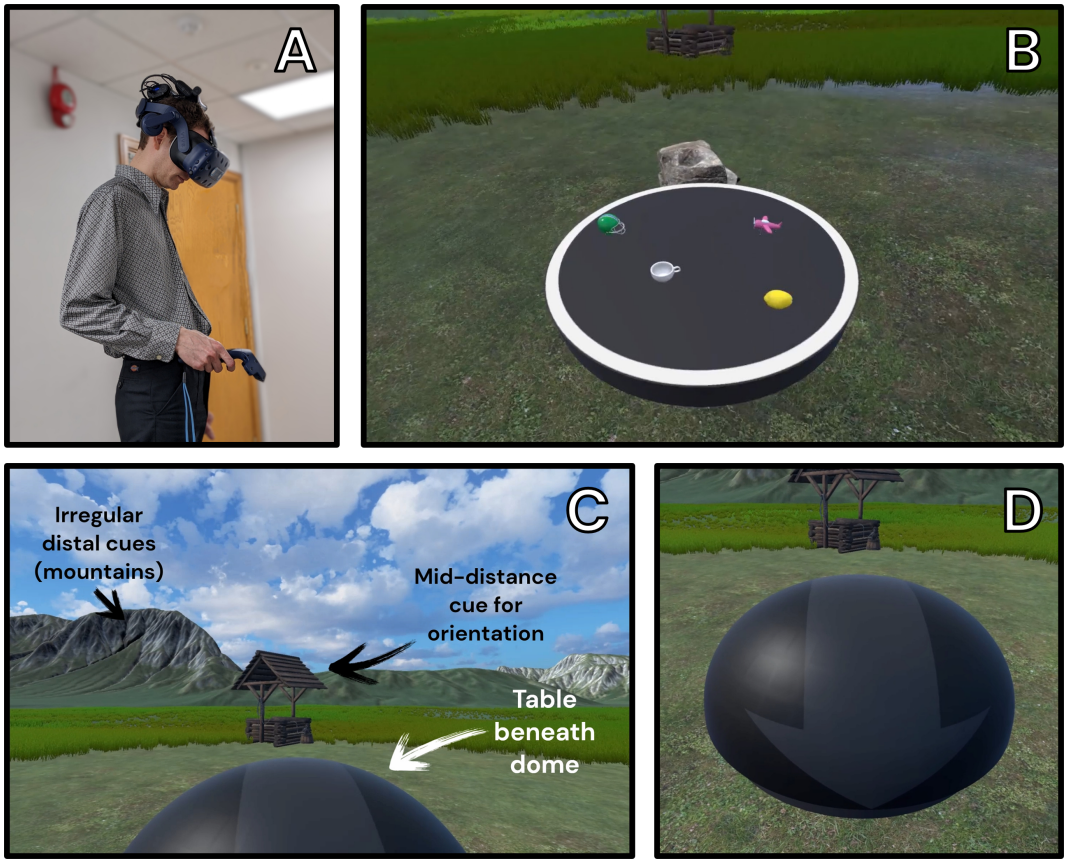


Figure 2.1. Stills of the equipment and virtual environment. (A) The HTC Vive Pro Eye with wireless adapter and controller (photograph included with permission). (B) View of four objects at encoding from a participant's view. Participants had 7 seconds to memorise the arrangement of objects. For a demonstration of exemplary trials, please see here: youtu.be/Wyw3wo7WMMI. (C) View of the wider virtual environment. For a 360° view, see here: youtu.be/IC0bgC_54. (D) The table was occluded by a dome between viewings.

2.2 Task design and trial structure

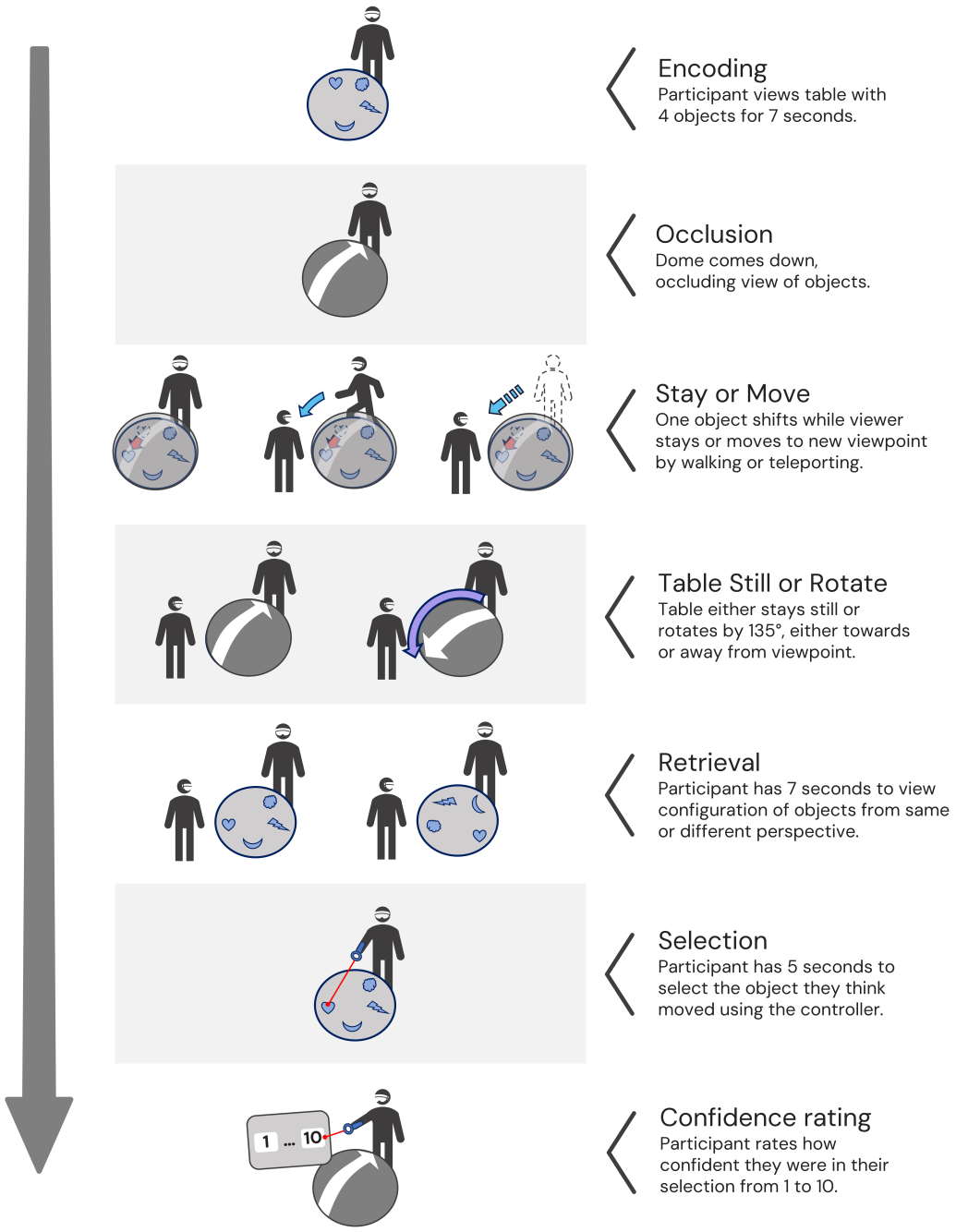


Figure 2.2. Schematic of trial structure. Six possible conditions were defined by combinations of participant movement (Stay, Walk, Teleport) and table rotation (Still, Rotate) between encoding and retrieval phases. Demonstration also available in video format youtu.be/Wyw3wo7WMMI.

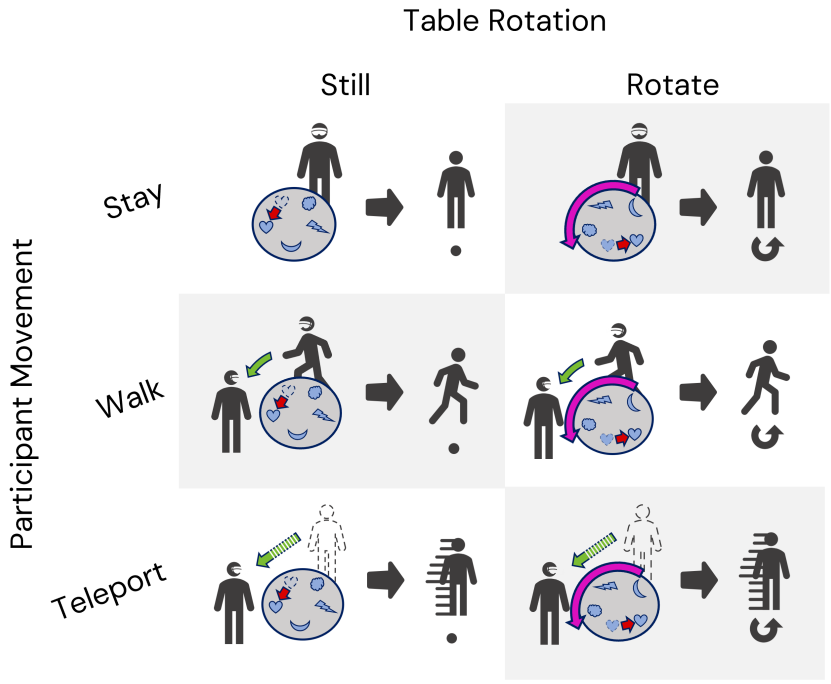


Figure 2.3. Conditions organised by table and participant movement with condition-icon key. Small icons will be used throughout the thesis to indicate conditions.

Participants were asked to detect which object had moved within an arrangement of four or five objects on a circular table between a first (encoding) and second (retrieval) viewing period, each lasting 7 seconds. They had to select which object they thought had moved, before rating their confidence in their choice on a scale from 1-10, with 1 being 'not at all confident' and 10 being 'extremely confident'.

First and second viewing periods were visually distinguishable by the colour of the table border. This was included as a visual reminder of the phase of the task, added following early piloting. After the second viewing, the table border changed colour again, indicating that it was time to select which object moved. This was done by pointing and clicking with the controller. Another 5 seconds was allowed here before the trial timed out and the occlusion dome came down (see Figure 2.2 and youtu.be/Wyw3wo7WMMI for a task demonstration).

In each trial, participants first viewed the array of objects from the starting viewpoint. The dome then occluded their view of the objects. Before the second viewing, they either stayed at the same viewpoint or moved to another viewpoint 135° around the centre of the table. This angular shift was chosen based on previous literature (Heywood-Everett et al., 2022) as it likely required a higher degree of allocentric spatial processing compared to smaller angles that can be partially solved using egocentric views (Burgess et al., 2004). It was also important to avoid angles that could allow idiosyncratic or non-spatial strategies, such as 90 or 180° (Mou and McNamara, 2002).

Trials that required participants to shift to a different viewpoint were split into two types of movement: walking and teleporting. This latter movement involved the participants instantly transposing their position within the virtual environment to the desired viewpoint with the click of a button. Specifically, participant movement was virtually transposed to the new position after they pointed to an enlarged viewpoint marker at the desired location and clicked the trigger on the controller. Participant rotation was not shifted, so they had to turn towards the table after teleporting. This also provided physical movement to control against the effect of self-motion in trials that required walking.

During half of the trials, the table rotated between first and second viewings by 135°. For trials where participants had moved between viewings, this meant the configuration of objects on the table was viewed from the same perspective at first and second viewings, but both the participant and the table had shifted with respect to external cues. For trials where the viewpoint was the same at first and second viewings, the table rotating meant the participants were viewing the configuration from a different perspective at the second viewing.

Accordingly, trials varied by both participant movement and table rotation. Participants either stayed in the first viewing position (‘Stay’), or moved to the shifted viewpoint by walking (‘Walk’) or teleporting (‘Teleport’). Additionally, the table either stayed still (‘Still’) or rotated (‘Rotate’), creating six different possible

trial conditions named after the change between viewings:


Stay-Still, the participant Stays and the table is Still: no change;


Stay-Rotate, only the table rotates away from the participant;


Walk-Still, the participant walks to the second viewing;


Walk-Rotate, the participant walks and the table rotates;


Teleport-Still, the participant teleports to the second viewing;


Teleport-Rotate, the participant teleports and the table rotates.

Figure 2.3 shows these conditions organised by participant and table movement. Note that Walk and Teleport combined will sometimes be referred to as Move (i.e. Move-Rotate or Move-Still trials).

Each participant completed 9 of each of the six conditions for a total of 54 trials. These were split into three sessions of 18 trials with three of each condition randomly ordered. Participants had a short break between each session. The rotation of the participant or viewpoint was counterbalanced to include include number of clockwise and anticlockwise rotations, which have shown to have an effect on memory in ageing (Castillo-Escamilla et al., 2022) but were not the target for hypotheses.

2.3 Objects

Configuration objects were a counterbalanced sample of four from 21 easily-recognisable and distinguisable three-dimensional models (shown in Section 2.5 in the next chapter), many selected from OpenVirtualObjects (Tromp et al., 2020) and some from previous experiments (Castegnaro, 2021, ch. 4, p. 149).

Chapter 3 discusses the choice of number of objects for the final version of the task.

Objects were contained within invisible spheres with 8 cm radius, equivalent to a 3-dimensional visual area of interest (AOI) for that object (see Figure 2.4). AOIs were larger than the bounds of the object's visible shape. This was to account for gaze points at the edge of the visible borders of objects, which may be erroneously detected as a gaze point on the table or an object behind it. The size of these AOIs was systemtically determined during development, as described in Chapter 3.

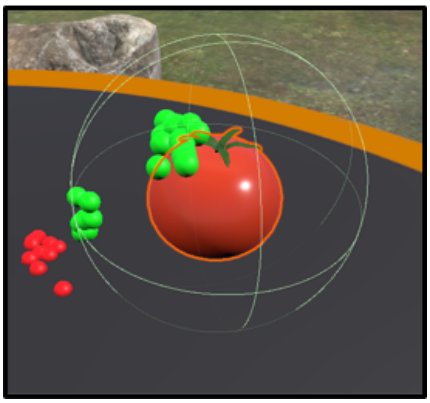


Figure 2.4. Visualisation of an area of interest (AOI) sphere surrounding an object, with gaze points registered there (green) and on the table nearby (red).

In each trial, four objects were placed on the table with contrasting colours and shape, without repetition within a trial or between consecutive trials. Object locations were generated with a pseudo-random placement algorithm. This first placed one object in a random position on the table, the next object on another random position on the table at least 28 cm away, and so on until all objects were placed. After placing all four objects, one object was randomly selected and moved to a pseudo-random location, maintaining the minimum 28cm between objects. This second configuration was then saved and the two configurations were shuffled to form a configuration pair.

Object shift distance was counterbalanced by condition by one of three

distances: 40±1cm, 55±1cm or 70±1cm. This allowed for a range of different types of configurations, and reduced the predictability of the object shift.

Configurations were discarded if they formed a regular shape (approximately equal angles or distances between objects in the convex hull) or if two or more objects were in line from the perspective of either viewpoint. Although these helped avoid regular shapes that were too easy to remember, the effect of the shape of the configuration on change detection is unknown. Therefore, one configuration pair was generated for each of the 54 trials, and configuration pairs were counterbalanced across participants and conditions to reduce any confounding effect of spatial configuration.

2.4 Procedure

Before beginning the experiment, participants were asked to name the objects from a printout (Figure 2.5). This was to avoid any distracting confusion over the identity of an object during the task.

Participants first underwent a training session of two guided trials at the very start and sixe further trials covering all conditions. After the training session, participants went through 13 practice trials. This included two trials per condition, plus an extra no change trial. For healthy control groups, if participants failed 2 out of 3 of the no change practice trials (the easiest condition), the practice was automatically cut short for the researcher to check the participant's understanding. The practice block was run again until the participant passed this criterion, up to a maximum of three times, which was never reached.

Participants then completed the experimental protocol (all three blocks) with five minute breaks between them. The whole procedure lasted approximately 90 minutes.

Figure 2.5. Final selection of objects used in the task.

2.5 Data collection

For each viewing period, data were recorded per frame for world-space gaze point, pupil diameter, eye openness, eye position, eye direction, headset position & rotation and controller position & rotation.

Object selection, confidence and trial parameters were recorded at the end of each trial. The locations of all visible objects in the environment were logged at the start of the session and after each change in position.

Data from the VR task were saved to raw data files, which were formatted, checked for validity and uploaded to a relational database ("PostgreSQL", 2022).

2.6 Data management, processing and engineering

Eye tracking data and the measures derived from the task were organised into different levels of nested granularity, visualised in Figure 2.6. The most fine-grained level was per frame, where eye movement data and geometric information on virtual objects were recorded. For eye-movements, frame-level data was used to calculate event-level measures, which in this case was split into fixations and saccades as described later in this section. These were used to compute viewing- and trial-level measures that were further divided into different conditions or combined on a per-participant, per-condition or per-group basis.

At the frame level, eye tracking data were checked for tracking loss. Periods

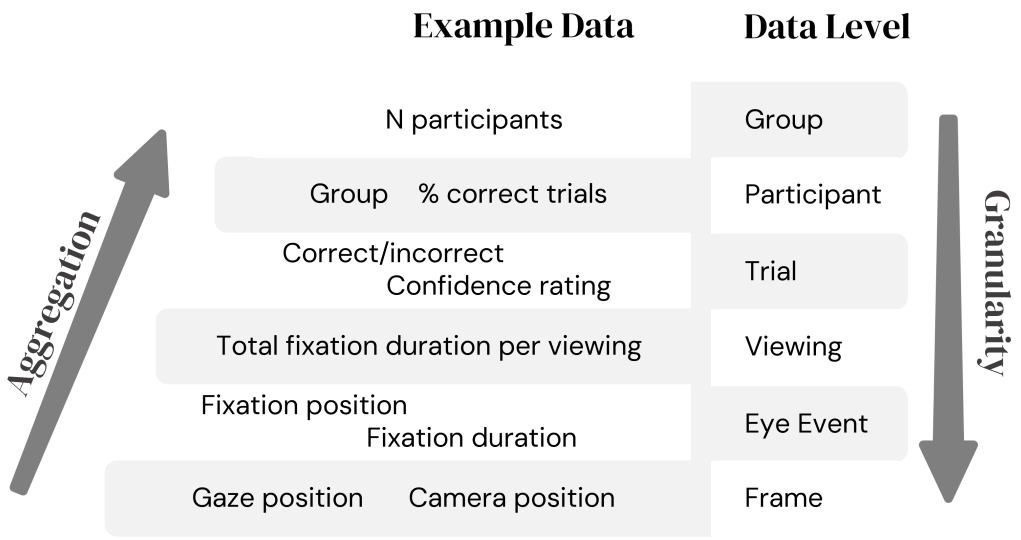


Figure 2.6. Diagram representing levels of granularity in the dataset. These levels will be referred to throughout the thesis.

of tracking loss were identified using built-in eye openness and pupil diameter parameters. These indicated whether an eye trace was validly recorded at each frame. Periods of tracking loss $<$100ms were considered technical malfunctions and were linearly interpolated (Ghose et al., 2020). For gaze point data, periods of tracking loss between 100ms and 1000ms were interpolated only if the distance before and after the tracking loss was less than 8cm (radius of an AOI sphere). This maximised the proportion of available data for estimating fixation times on AOIs by assuming that these periods of tracking loss were during a fixation. All other periods of tracking loss were marked as missing data and linearly interpolated for filtering, but not included in further analysis.

Frame-level data were also filtered and smoothed to improve signal-to-noise ratio, including eye and head position data. Filtering approaches followed previously documented procedures (Das et al., 1996; Diaz et al., 2013). All frame-level signals were up-sampled to 1000 Hz using linear interpolation before filtering with a first-order low-pass Butterworth filter with 30 Hz cutoff frequency. Eye data were then filtered with a 5-point (per 120 Hz frame) median filter and

5-point moving average filter. Camera position and rotation were filtered with a 7-point median filter and 5-point moving average filter. Derivatives of gaze and camera signals were calculated from these signals (e.g. velocity).

Frame-level data were used to classify two distinct functional and oculomotor events: fixations and saccades. The former are periods of maintained, localised gaze used for visual information gathering, accounting for between 70% and 90% of gaze time (Bogartz and Staub, 2012; Einhäuser et al., 2006; Ford et al., 1959). Saccades, conversely, are the fast, ballistic eye movements between fixations, during which visual information is actually suppressed (Matin, 1974). Identification of fixations from saccades is a common step in processing eye-tracking data (Salvucci and Goldberg, 2000). The approach to this is detailed in the next chapter, Section 3.2.5.

Smaller oculomotor events were not identified due to the low precision of the eye-tracker, including post-saccadic oscillations and microsaccades (Holmqvist and Blignaut, 2020). Additionally, smooth-pursuit eye movements were not examined here because all AOIs were static during viewing periods. Finally, pupillometric measures were not analysed as they were not relevant to hypotheses.

2.6.1 Eye movement measures

After partitioning eye movement data into fixations and saccades, many different measures can be extracted to summarise or describe viewing behaviour. Typically, an appropriate time period is chosen and the number or duration of fixations on AOIs are calculated as straightforward memory-related eye movement derivatives. I adopted this approach as the most basic method of analysing eye movements, calculated from the standard 7s viewing period at encoding and retrieval. These viewing-level measures were used to compare viewings within and across participants.

However, summarising eye measures at the viewing level fails to account

for most temporal or sequential information. Although researchers may extract measures derived from the temporal dimension, such as the time before the first fixation on a key AOI. However, this still ignores potentially insightful temporal dynamics of viewing behaviour within viewing periods.

There are several analytical approaches available to account for dynamic time data. One approach involves calculating the probability of fixating on an AOI over time and fitting models to the resulting time-series data (Oleson et al., 2017). This approach allows for time-dependent measurements of viewing behaviour based on an averaged viewing period, including the change in gaze behaviour. An implementation of this approach is described in chapter 5.

While this type of analysis can search for changes in viewing patterns over time, it does not easily account for the viewing behaviour between AOIs. A number of studies have focused on studying fixations based on the transition patterns between AOIs, sometimes referred to as the 'scan-path' (Anderson et al., 2015; Noton and Stark, 1971; Yarbus, 1967). This class of measurement requires calculation of the transitions between AOIs, usually combining fixations within AOIs into one 'external' fixation.

There has been some debate on the best way to analyse scan-paths (Hayes et al., 2011; Wollstadt et al., 2021). One popular approach involves modelling them as Markov chains because each AOI can be represented as states in a fixation sequence, with transitions between them. Transition probabilities can be calculated between AOIs and information theoretic measures such as Shannon Entropy can be calculated per scan-path (Krejtz et al., 2015; Krejtz et al., 2014; Ebeid and Gwizdka, 2018; Lee et al., 2022). Indeed, scan-path entropy has been shown to be higher in participants with medial temporal lobe damage (Lucas et al., 2018) and MCI (Coco et al., 2021) and was therefore included as an outcome measure in this study.

However, this approach still ignores most temporal information, because a Markov chain assumes that each state is dependent only on the previous one,

whereas evidence supports eye movement planning beyond the next fixation (Hoppe and Rothkopf, 2019). More advanced techniques have been developed to account for this, but they either sacrifice ease of interpretation (Hayes et al., 2011) or comparability across differing scan-paths (Wollstadt et al., 2021). For this latter issue, a range of scan-path comparison statistics have been developed to measure the similarity between one fixation sequence and another. For example, the Levenshtein distance or string-edit method measures the distance to edit one string of symbols into another, with a cost for insertion, deletions and substitutions (Mathot et al., 2012). This algorithm was adapted for 'MultiMatch', a set of measures that quantify the similarity between a pair of scan-paths based on the sequence of fixation durations, distances, or vectors (Dewhurst et al., 2012; Jarodzka et al., 2010).

When using a comparative measure to quantify the similarity between two scan-paths, one loses the advantage of a single statistic per scan-path that can be compared between trials or participants. In other words, similarity metrics like Levenshtein distance and MultiMatch are relative measures, requiring two scan-paths for calculation. Although a limitation in some cases, this is a useful approach for paradigms that involve two or more viewing periods per trial such as this one. Specifically, a key eye movement analysis involved the approximation of a gaze reinstatement effect (see Section 1.4) by comparing scan-paths at encoding and retrieval viewing periods for each trial. However, some design features of the current paradigm created complications for quantifying similarities between scan-paths at the first and second viewings. Firstly, the scene changed between the two viewings: moving one object not only moved the AOI, it also created an extra AOI: the previous position of the moved object. Furthermore, viewpoint shifts and table rotations changed the world-space and egocentric positions of the objects.

To partially accommodate these issues, the MultiMatch 'shape' similarity metric was used to quantify a gaze reinstatement effect, but only in trials without

table rotation. This meant that a comparison between the world-space vector sequences at encoding and retrieval were made to estimate gaze reinstatement. An analysis of an egocentric gaze reinstatement effect in Move-Rotate trials was not performed for this thesis.

In addition to comparing scan-paths within trials, similarity metrics provide a potential means of comparing scan-paths between participants on similar stimuli. Accordingly, a data-driven methodology for analysing scan-path differences at encoding was developed by combining MultiMatch with a clustering algorithm. This did not contribute to the results of the thesis, but can be found in Appendix A.

Chapter 3

Feasibility of the paradigm for testing younger and older healthy adults

Abstract

The feasibility of the new immersive virtual reality task was tested in healthy younger and older adults.

Methodology. Four- and five-object versions of the task were assessed for performance in younger and older adults with a view to prevent any floor effects. Eye tracking precision and data collection feasibility were tested, including calibration of areas of interest (AOIs) around objects. A bespoke fixation identification procedure was developed and tuned against a published method. Eye movement associations with task performance were examined to test for task modulation of fixation patterns. Usability and acceptability of the task procedure was assessed by brief semi-structured interview.

Summary of findings. All participants found the task acceptable and usable, although several participants commented on its difficulty. The five-object version of the task was found to be too difficult for older adults specifically, with several participants performing at or below chance. AOI sizes required adjustment after assessing erroneous fixations. Eye tracking precision was low for some visual field targets. A new 'GazeCollide' fixation identification algorithm was developed, showing improved data processing speed over the most comparable method. Participants had significantly greater fixation time on the moved object in correct trials, supporting task-relevant effects of eye movements.

Conclusions. A four-object version of the task was used for further data collection. Precision of the integrated eye-tracker was likely too low for saccadic measures; eye metrics based on fixations were employed for analysis of task results from a larger sample.

3.1 Introduction

Before testing any scientific hypotheses, the feasibility of the new task was tested. The term *feasibility study* was used in the broad sense here: in determining whether it was possible or reasonable to run a larger-scale research project on the paradigm in question. Indeed, the main purpose of a feasibility study in research is to ascertain whether a paradigm *can* work to test hypotheses, rather than test the hypotheses themselves (Bowen et al., 2009). As a result, methodologies can vary greatly according to the specifics of the study design. However, they commonly include an assessment of user experience, often allowing participants to suggest improvements to the paradigm. For this study, I combined this user-focused assessment with some task-specific feasibility tests of technical functioning, task difficulty, and eye-tracking data processing.

For assessing task difficulty, the most important population to test for feasibility were age-matched control participants. If these participants found the task difficult to perform or use, it was likely due to age-related factors, or the design of the task itself, which I aimed to dissociate from Alzheimer's-related cognitive and behavioural changes. Any potential floor effects could obscure differences between healthy and patient groups when aiming to find a measure that discriminated between them. An appropriate task difficulty therefore required older adult participants to perform above chance—the probability of a correct answer if they were randomly guessing.

Ideally, the feasibility of the task in older adults with memory impairments would have also been tested, if not to assess their memory performance, then to check usability and acceptability in a population that may have had different needs than their unimpaired counterparts. Unfortunately, participants with mild cognitive impairment (MCI) were unable to be recruited when the task was ready for feasibility testing due to delays related to the COVID-19 pandemic. However, I have included qualitative results of the first five patient participants, collected at a later date for the main study results described in Chapter 4. The details

of the task itself could not be changed after these participants were tested, but their experiences were still documented.

In addition, a younger adult group was recruited to test the feasibility of examining age-related differences and replicating previous findings (Burgess et al., 2004; Simons and Wang, 1998). One consideration for making the task appropriate for older adults was a potential ceiling effect in this younger group: the task could become too easy for them. It was preferable to ensure both age groups were not performing at floor or ceiling, but because I was most focused on calibrating the difficulty to an age-matched control group for patient comparison, it was a priority to keep the older group above chance.

As this was the first use of this eye-tracking equipment for this paradigm, the feasibility of collecting eye movements during the task was important to confirm. Eye tracking allows for rich data collection, but the data can be low quality and require several pre-processing steps (see Section 2.6). For example, to make sure the eye data was usable for testing scientific hypotheses, I tested the technical feasibility of calibrating the eye tracking consistently, the precision and accuracy of the eye tracker, and early hypothesis-related results based on eye movement measures.

An important processing step for the eye-tracking data was partitioning gaze samples into fixations and saccades. A bespoke algorithm was developed for this purpose. Therefore, a key feasibility step was to validate its use and calibrate its parameter thresholds. Eye movement comparisons between groups relied on this algorithm in later chapters. Therefore, to assess whether the algorithm was fit for use, key fixation-derived measures were tested for task-relevant effects. For example, an association between fixations on the moved object and task performance was expected even with a small, feasibility sample. This step provided a check that eye-tracking data were potentially useful or appropriate for testing hypotheses.

3.2 Methods

3.2.1 Task Design

The task structure was identical to the design detailed in the General Methods Section 2.2, except that some participants were given a version of the task with five objects per configuration, and some with four. This was the main method of calibrating the task difficulty for older adults. A five-object task would be more consistent with earlier versions of the viewpoint-shifting paradigm (Burgess et al., 2004; Mou and McNamara, 2002; Simons and Wang, 1998). However, an easier four-object version was ultimately more appropriate for older adults, who had not been tested on the paradigm before.

3.2.2 Participants

Thirteen healthy younger participants (7 female, mean age 25.1, age range 20-33) and eleven healthy older adults (6 female, mean age 73.1, age range 61-80) participated in the study after giving their informed consent. The study was approved by the UCL Research Ethics Committee (SHaPS-2018-JK-027). Only participants without the following were recruited:

- A visual impairment that could not be corrected-to-normal, including colour visual deficiency (colour-blindness);
- Significant mobility problems or injuries that may cause problems with the task;
- Significant mental health disorders;
- Difficulty with verbal or written instructions due to communication needs or language barriers;
- A diagnosis of dementia;
- Other neurological conditions including cerebrovascular disease.

In addition, five participants with mild cognitive impairment (MCI) were asked qualitative feasibility questions, but did not contribute to the dataset for other areas of feasibility testing. These participants were recruited with the same eligibility criteria aside from allowing for the diagnosis itself. The details of the diagnostic criteria for MCI can be found in Chapter 4 (Section 4.2.1).

3.2.3 Qualitative feasibility

A brief, semi-structured interview was used to collect data on usability and acceptability, as well as researcher observation. Participants were all asked the following, with encouragements to expand upon answers where appropriate:

1. how clear the training instructions were;
2. how easy or difficult they found the task to navigate (ignoring perceptions of performance);
3. their understanding of the task after practice trials;
4. whether they found anything confusing about the task;
5. how they found the visibility of the objects;
6. their overall experience of the task.

Answers were recorded in writing for content analysis. Additionally, the researcher observed the following:

1. participants' ability to navigate the task,
2. participants' tolerance of the task and VR in general, and
3. whether any safety risks arose.

3.2.4 Technical feasibility and data quality

3.2.4.1 Eye tracking calibration consistency

Functioning of the manufacturer's built-in eye tracking calibration was noted for each participant by the researcher. The only details of calibration results for

developers was success or failure. If the calibration failed, it was attempted up to twice more, including after a system reboot.

3.2.4.2 Precision and accuracy

Measuring precision and accuracy was important for comparing eye tracking results across studies, determining the quality of the data, and assessing the suitability of the eye movement analysis. For example, if eye tracking was too inaccurate or imprecise, then it would be inappropriate to examine measures such as intra-fixation gaze patterns.

The precision and accuracy of the eye tracking were assessed using GazeMetrics (Adhanom et al., 2020). This open-source software was designed for immersive virtual reality (iVR) headsets with integrated eye-tracking. These data were useful to collect for the HTC Vive Pro Eye system because, as mentioned, the manufacturer's calibration software did not allow developers access to detailed results.

GazeMetrics appeared as a calibration programme, providing a series of dot targets to fixate on, locked to the participant's view (Figure 3.2A). The accuracy measurement was calculated as the difference between the reported fixation location and the intended target. The spatial precision metric quantified the variability in gaze measurements by calculating the root mean square (RMS) of the angular differences between eye vector samples (equations for accuracy and precision are shown in equations 1 and 2 in Adhanom et al., 2020).

GazeMetrics was tested on participants after in-built calibration and before the task. Accuracy and precision were recorded per fixation target on raw eye-tracking data; no pre-processing steps described in Section 2.6 were performed prior to this.

3.2.5 Fixation identification

A custom eye-event algorithm was written to classify fixations from saccades. Traditional identification of fixations from saccades has involved manual labelling

by one or more human expert coders. This usually relies on a set of rules and at least two coders, who must demonstrate a minimum level of agreement. However, sourcing expert coders with the availability to code hundreds of viewing periods is a significant limitation of manual coding, and precluded it from this study. Moreover, manual approaches allow for more uncontrolled subjectivity, and agreement between coders can vary significantly, affecting event-level features such as the number of fixations (Hooge et al., 2018). This has lead some to conclude that it should not be considered the gold-standard fixation identification technique (Hooge et al., 2018).

Alternatively, identifying fixations from saccades is commonly performed by one of several computational algorithms. The problem lends itself well to an algorithmic approach because these eye-events are often defined by changes in derivatives of eye position such as dispersion, velocity, or acceleration. However, there is no clear consensus on which is the most effective or reliable fixation identification algorithm (as reviewed in Andersson et al., 2017). The lack of agreement is likely because the most widely used methods are based on thresholds of velocity or spatial dispersion, which may not generalize across eye-tracking equipment, study design, or pre-processing steps. Efforts to create generalised algorithms have been published (e.g. Zemblys et al., 2019), but this can come at the expense of simplicity or data processing speed.

Choosing the appropriate fixation identification algorithm for VR-based eye tracking requires some unique considerations compared to traditional eye tracking methods, and also opens up unique opportunities. In VR environments, the user is interacting with a 3D world, and therefore their head movement must be taken into account in the eye tracking calculations. These calculations are integrated into the VR development software, but traditional fixation identification algorithms must still be adapted to account for free head movement during a fixation period.

Very few algorithms have been developed or adapted for use with VR-based

eye tracking. One study adapted the Identification by Dispersion Threshold algorithm (Salvucci and Goldberg, 2000) for use with VR (VR-IDT) using the HTC Vive Pro Eye (Llanes-Jurado et al., 2020. The algorithm searches for fixations by calculating the angular eye shift frame-to-frame and applying a dispersion threshold with a moving time window of a minimum length. To adapt this approach to iVR, if a fixation is identified based on dispersion and time thresholds, head movement is averaged for the fixation.

To the best of my knowledge, VR-IDT is the only published method for eye-event classification with the same equipment as this study (although see Chen and Hou, 2022 for an unpublished approach with many similarities). Unfortunately, the algorithm is relatively slow to label data; VR-IDT takes on average 1.4 seconds to label a 7 second viewing period from this study (Figure 3.3). The anticipated dataset for the diagnostic study was estimated to constitute $>$10,000 viewing periods, which would take $>$4 hours to process fixations alone. The iterative nature of data analysis would likely require re-running fixation processing several times, adding a considerable time burden on data processing.

3.2.5.1 GazeCollide

Although several alternative methods exist that could be adapted to the VR system (Salvucci and Goldberg, 2000), a custom fixation identification algorithm was developed to take advantage of the integration of eye-tracking and visual stimuli in a single system. Using the SRanipal software development kit implemented in the HTC Vive Pro Eye, the foveated object can be registered at runtime. SRanipal interacts with the Unity game engine to automatically project a line in 3D space using the known location and orientation of the eyes as the origin and direction of the line vector. As a developer, it is then relatively simple to register where this gaze line 'collides' with world-based objects. For a task such as this one, where AOIs are well-defined and spaced apart (i.e. the objects, with the exception of the Table, discussed in Section 3.2.6), fixations can be mostly

demarcated by which object is registered per frame. This is because they essentially provide stimuli-driven dispersion thresholds. Of course, this relies on pre-defined AOI boundaries, which were determined using a custom methodology detailed in Section 3.2.6 of this chapter. Additionally, a velocity threshold was applied to detect saccades within AOIs, which was particularly important for consecutive fixations on the table itself. This was calculated by the angular velocity in the eye direction, rather than the world-space point-to-point velocity, which is not appropriate for 3D environments. A formalisation of this algorithm, named 'GazeCollide' can be found in Algorithm 1.

Algorithm 1: GazeCollide algorithm

Data: *Gaze Array*: array of gaze point tuples (t, o, g_x, g_y g_z):
timestamp t; gaze object o; and gaze collision point g x, y and z.
Parameters: Max. Velocity v_{max}, Min. Duration d_{min}
Result: *Fixation labels:* array $f_{1,2,...,t}$ with $1/0$ for fixation/saccade.
Result: *Fixation start labels:* array $s_{1,2,...,t}$ with 1 marking the start of a fixation and 0 not.
Result: *Fixation end labels:* array $e_{1,2,...,t}$.

$f_{1,2,...,t} \leftarrow 1$, assign 1 to all f;
for *all* $T_2, T_3, ..., T_t$*, vectorized* **do**
    Calculate angular velocity v_t between T_t and T_{t-1};
    if $v_t > v_{max}$ **then**
        $f_t \leftarrow 0$, mark as saccade; `/* velocity threshold */`
    if $o_t \neq o_{t-1}$ ***and*** $f_t = 1$ **then**
        $s_t \leftarrow 1$; `/* fixation start */`
    if $o_t \neq o_{t+1}$ ***and*** $f_t = 1$ **then**
        $e_t \leftarrow 1$; `/* fixation end */`
Assign 1 to all f between where $s = 1$ and $e = 1$;
Combine consecutive fixations if separated by one gaze point;
for *all consecutive points where* $f = 1$ **do**
    if *duration of consecutive points* $d < d_{min}$ **then**
        $f \leftarrow 0$

GazeCollide was designed with processing speed in mind where possible. However, another reason for developing this new algorithm was its suitability to

the hypotheses of the study. My main eye-tracking hypotheses related to which AOIs participants fixated on and for how long. GazeCollide was inherently based on this information, while filtering out gaze samples that were too fast to be fixations. In contrast, VR-IDT is based on the angular movement of the pupils; the gaze object must be identified after fixation labelling and one fixation may cross multiple objects, requiring further heuristics and post-processing to label a fixation as being on one AOI instead of another.

However, it should also be acknowledged that GazeCollide is a bespoke method and may not generalise well to other studies. Indeed, the paradigm for this study varied from the VR-IDT study (Llanes-Jurado et al., 2020) in several key ways, one being that the main AOIs—the configuration objects—were deliberately placed a minimum distance from each other. Although viewers could look at the table near or between objects, the most salient visual stimuli were spaced apart by 28 cm, unlike Llanes-Jurado and colleagues' study, avoiding some ambiguity in the foveated AOI.

3.2.5.2 Calibrating GazeCollide thresholds

As discussed, there is no real gold standard by which to validate new eye-event detection algorithms (Hooge et al., 2018). The best option is to validate against manual labelling of established datasets (such as in Zemblys et al., 2019). However, reliable expert labelling was not feasible for this study due to time constraints, and validation of GazeCollide on an existing dataset was not possible because of differences in equipment and the algorithm's need for distinct, pre-registered AOIs. Despite my previous reasoning against its use, the most relevant eye-event algorithm to compare GazeCollide to is VR-IDT, which has been calibrated for use in experimental paradigms with the same equipment (Llanes-Jurado et al., 2020). Accordingly, GazeCollide was compared against VR-IDT in two key ways. Firstly, in fixation labelling speed: both algorithms were timed on classification of fixations and saccades for 100 random trials (2 viewing pe-

riods per trial) from the feasibility sample and compared for differences using a paired-samples *t*-test with a one-tailed alternative hypothesis that GazeCollide would run significantly faster than VR-IDT.

The second way GazeCollide was compared to VR-IDT was in tuning the velocity threshold of the former. Any new fixation algorithm should identify a proportion of fixation samples consistent with eye-movement literature, which suggests between 70% and 95% for adult eye-tracking studies (Bogartz and Staub, 2012; Einhäuser et al., 2006; Ford et al., 1959). Indeed, the optimised thresholds for the VR-IDT algorithm resulted in ∼90% fixation samples in their original study (100ms time window, 1.6° dispersion threshold). To ensure GazeCollide captured an appropriate proportion of fixation samples, this was calculated per viewing for a range of velocity thresholds and matched to the same measure from VR-IDT using author-optimised thresholds. This was achieved by calculating the sum of square differences between the proportion of fixation samples identified by each GazeCollide velocity threshold and the optimised VR-IDT algorithm, as described in equations 3.1-3.3.

$$\vartheta^* = \underset{\vartheta \in \theta}{\operatorname{argmin}} \Big(S_\vartheta \Big) \tag{3.1}$$

$$S = \sum_{j=1}^{v} \Big(P_{dj} - P_{g_j} \Big)^2 \tag{3.2}$$

$$P = \frac{1}{\sum_{k=1}^{n} T_k} \times \sum_{k=1}^{n} T_k F_k \tag{3.3}$$

In words, the proportion of fixation samples were calculated for each viewing period using equation 3.3 by multiplying each gaze point duration in T by it's corresponding fixation label in F as determined by the fixation algorithm. A fixation label was 1 for fixations and 0 for saccades, so multiplication would result in the sum of fixation gaze point durations. This sum was then divided by the total gaze duration for that viewing period. Then, for each velocity threshold

ϑ in θ, P for GazeCollide algorithm g was subtracted from the VR-IDT algorithm d for each j in viewing periods v. The minimum sum of squares was then used to select the optimum velocity threshold ϑ^*.

Previous velocity-based algorithm thresholds range from $<$20°/s to 100°/s (Salvucci and Goldberg, 2000; Zelinsky and Loschky, 2005; Spieler et al., 2006; Llanes-Jurado et al., 2020; Nárai et al., 2021). Based on visual inspection of velocity-time charts (see Figure 3.4a for an example), I used a range of 20°/s to 90°/s in steps of 10°/s as parameters in GazeCollide for equation 3.2. A minimum fixation duration threshold of 100ms was applied to all algorithms, as this has commonly been found as a useful threshold for removing fixations that are too short for processing visual information and therefore unable to be fixations (Manor and Gordon, 2003). It is also an optimised parameter for VR-IDT (Llanes-Jurado et al., 2020).

To check that changing velocity thresholds would not make a difference to key outcome measures, each velocity threshold was used to calculate four eye-tracking metrics:

1. The proportion of fixations on the moved object, which was relevant to the main study hypotheses, and tested for changes to fixations on configuration objects.

2. The proportion of fixation time, or 'dwell proportion' on the moved object;

3. The proportion of fixations on the table. This was chosen because fixations on the table, while still relevant to hypotheses (detailed in Section 4.3.1 in the next chapter), were on less salient visual features of the environment (less to focus on) than the configuration objects and may therefore show different fixation patterns.

4. The dwell proportion on the table.

To test for main effect of algorithm-threshold pair on fixation-derived metrics, one-way Analyses of Variance (ANOVAs) were calculated per metric $(\alpha < .05)$. Further, to test for any group-dependent effect of varying the threshold, the interaction between velocity threshold and group in predicting each metric was calculated using a two-way mixed-effects ANOVA.

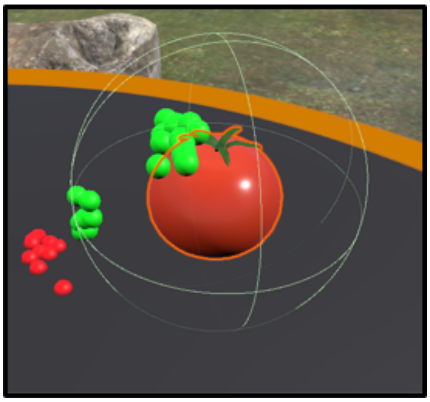


Figure 3.1. Visualisation of an area of interest (AOI) sphere surrounding an object (green wire frame), reproduced from Chapter 2. Green markers are gaze points registered to the object, red markers are gaze points on the table nearby.

3.2.6 Calibrating the size of AOI spheres

The second threshold to calibrate for GazeCollide was the spatial definition of each AOI. This was determined by the radius of the collision sphere for objects on the table (see Figure 2.4), and the outer mesh for all other objects (i.e., the visible borders of objects such as the table, the occlusion dome and external cues). For configuration objects, a radius too small would lead to more gaze points erroneously registered on the table. It was therefore important to choose a sensible size of these spheres. However, determining appropriate sizes required a different approach to calibrating the velocity threshold in GazeCollide; choosing a set of thresholds from the literature and calibrating against VR-IDT, or testing sensitivity to key outcome measures, was not appropriate here because changing the AOI size around configuration objects would not necessar-

ily change the fixation/saccade ratio. Rather, it would predominantly change the ratio of gaze points on the table to gaze points on configuration objects. Therefore, the size of AOI spheres was not an algorithm-specific threshold, but a task-relevant one.

To estimate an appropriate size of AOI spheres, a separate sub-task was used to ensure that gaze points intended for a configuration object were registered as such. Participants underwent the sub-task, named the Object Calibration phase, before starting the training trials. This required them to look at a selection of the objects, one at a time, at five predefined locations on the table (see Figure 3.6b). These locations were chosen to account for differences in precision across the field-of-view of the iVR head-mounted display (addressed in Section 3.3.1.2 of this chapter). AOI radii ranging from 4 cm to 12 cm in steps of 2 cm were chosen based on the size of the objects and distance between object positions. Smaller increments in radius were avoided to prevent over-fitting parameters to the feasibility sample.

Judging whether gaze points were registered correctly was estimated via the number of fixations registered on the table, as no fixation should be on the table when the participant was looking at the objects in the Object Calibration phase. AOI size was increased incrementally, and any radius that removed table fixations was recorded, after accounting for any fixations on the table beyond 12 cm (which were assumed to be not intended for the object). The positions on the table at which most table fixations were removed was also recorded. Note that this was done for all fixations from all participants, not within participants.

Visual inspection of gaze and fixation data was also required to validate whether table fixations were indeed erroneous. To this end, a Unity program was developed wherein visual renderings of participants' eye movements during the Object Calibration phase could be inspected (Figure 3.7). In brief, eye and head movement data were fed into a program that could 'replay' a rendering of the task as the participant experienced it but with visual representations of gaze

points and fixations overlaid.

3.2.7 Performance feasibility

As mentioned early in this chapter, the difficulty of the task was important to gauge for healthy older participants. Therefore, the effect of group and number of objects were tested on percentage of correct trials using a two-way between-subjects ANOVA. This was further split by condition to examine the performance in key conditions such as Walk-Still. Additionally, percentage of correct trials was compared to chance (20% correct with 5 objects; 25% with 4 objects) using one-tailed, one-sample *t*-tests for each condition-group pairing.

However, the sample size was under-powered for these statistical comparisons at an alpha level of 0.05. Statistical output of ANOVA calculations were reported in this context alongside visual inspection of bar charts.

To assess whether eye movements and memory performance were significantly associated based on this early data, fixation time on the moved object at retrieval was compared between correct and incorrect trials. This eye movement measure was central to scientific hypotheses of the wider study. A hierarchical mixed-effects linear model was run per age group to test for main effects of correctness, with trial data nested within participant groupings.

3.3 Results

3.3.1 Technical and data quality results

3.3.1.1 Eye tracking calibration

All eye tracking calibration failures were resolved prior to practice phases. This was only required for two participants, whose calibration failed on the first attempt only.

3.3.1.2 Precision and accuracy

GazeMetrics was tested on 11 participants (some data were lost due to a technical error). The average precision and accuracy of the eye tracking at different locations as measured by GazeMetrics is shown in figures 3.2B and 3.2C. These specifications are less accurate and precise than many industry-standard eye trackers (Clay et al., 2019). Note that the manufacturer states 0.5-1.1° angular error at target 1. However, target 1 had the lowest accuracy and precision (and highest median absolute deviation of each). This may have been because it was the first target presented and participants were still becoming accustomed to the task. Future tests of accuracy and precision may need to counterbalance the order of targets.

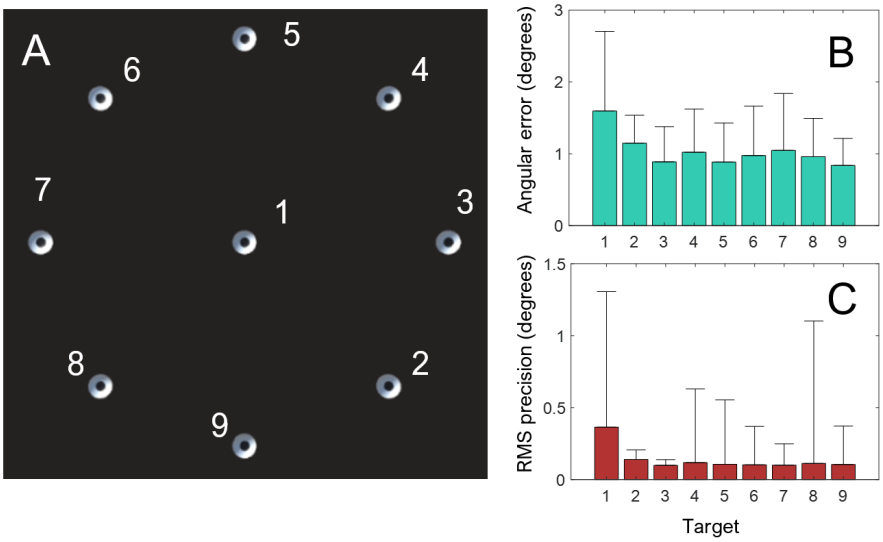


Figure 3.2. (A) GazeMetrics fixation targets, numbers added for reference in order of appearance. These targets were locked to the participant's view. (B and C) Median angular error of eye fixation, an inverse measure of accuracy, and median RMS, an inverse measure of precision, at each numbered target in A. Error bars are median positive deviations.

3.3.2 GazeColide calibration and comparison to VR-IDT

Figure 3.3 shows the difference in fixation labelling speed between GazeCollide and VR-IDT algorithms on a logarithmic scale. GazeCollide completed fixation labelling significantly faster than VR-IDT, $(t(199) = -8.07, p < .0001)$. The mean increase in speed was 10.7x that of VR-IDT. For 100 participants with 54 trials each, GazeCollide would label fixations in $\sim$25 minutes, whereas VR-IDT would take $\sim$4.2 hours.

To calibrate velocity thresholds for GazeCollide, Figure 3.4 shows the percentage of fixation time labelled by GazeCollide across a range of velocity thresholds in comparison to VR-IDT at the authors' optimised dispersion threshold of 1.6°. A velocity threshold of 60°/ s was determined as the optimal velocity threshold from the set of thresholds shown, such that eye movements above this threshold were automatically labelled as saccades.

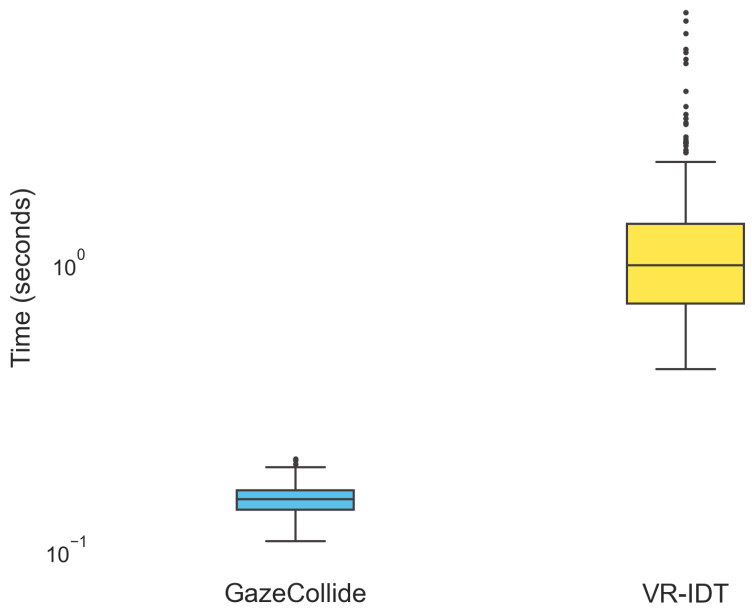


Figure 3.3. Comparison between GazeCollide and VR-IDT algorithms in processing time per viewing on a logarithmic scale. Taken from a sample of 200 viewings.

To test for an effect of algorithm and threshold on eye-movement metrics, Figure 3.5 shows the comparison between algorithm-threshold pairs on key measures. No significant main effect of algorithm-threshold pair was found on any of the four measures, $(p > .05)$, indicating that no evidence could be found for an effect of varying the fixation algorithm or velocity threshold on hypothesis-related eye-movement measures. Additionally, no significant effect of algorithm-threshold pair interacting with group was found in predicting any measures, $(p > .05)$, therefore no evidence was found for an effect of changing algorithm, or algorithm parameters, on the difference in hypothesis-related eye movement measures between groups.

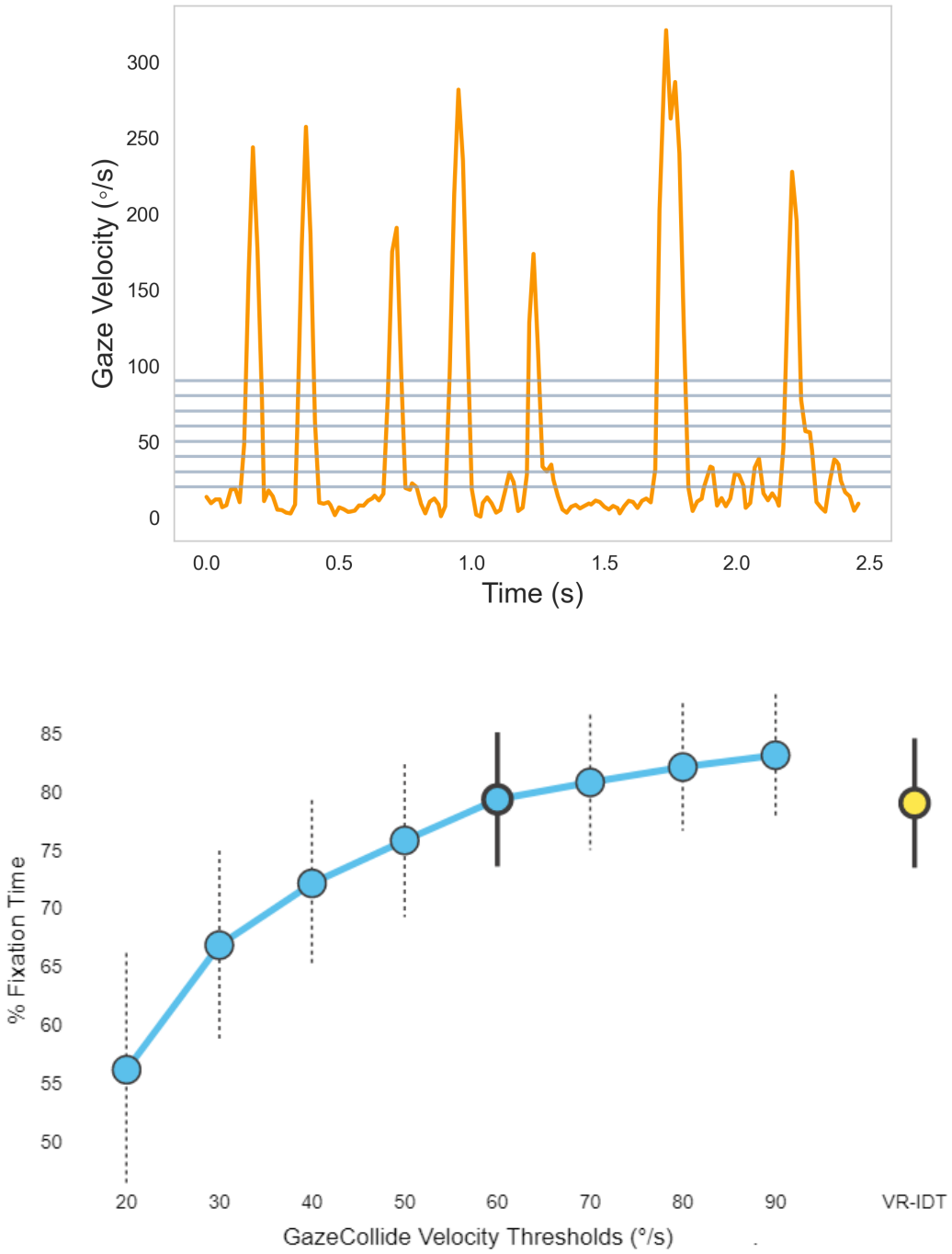


Figure 3.4. Velocity thresholds for calibrating GazeCollide. [Top] an example segment of gaze velocity data over time (orange line) with velocity thresholds superimposed (grey horizontal lines). [Bottom] Dot plots of mean and standard deviation percentage of fixation time as labelled by GazeCollide with different velocity thresholds. GazeCollide at a velocity threshold of 60°/s was the optimal match to VR-IDT with 1.6° dispersion threshold as determined by equations 3.1-3.3. This is represented by thicker, solid lines.

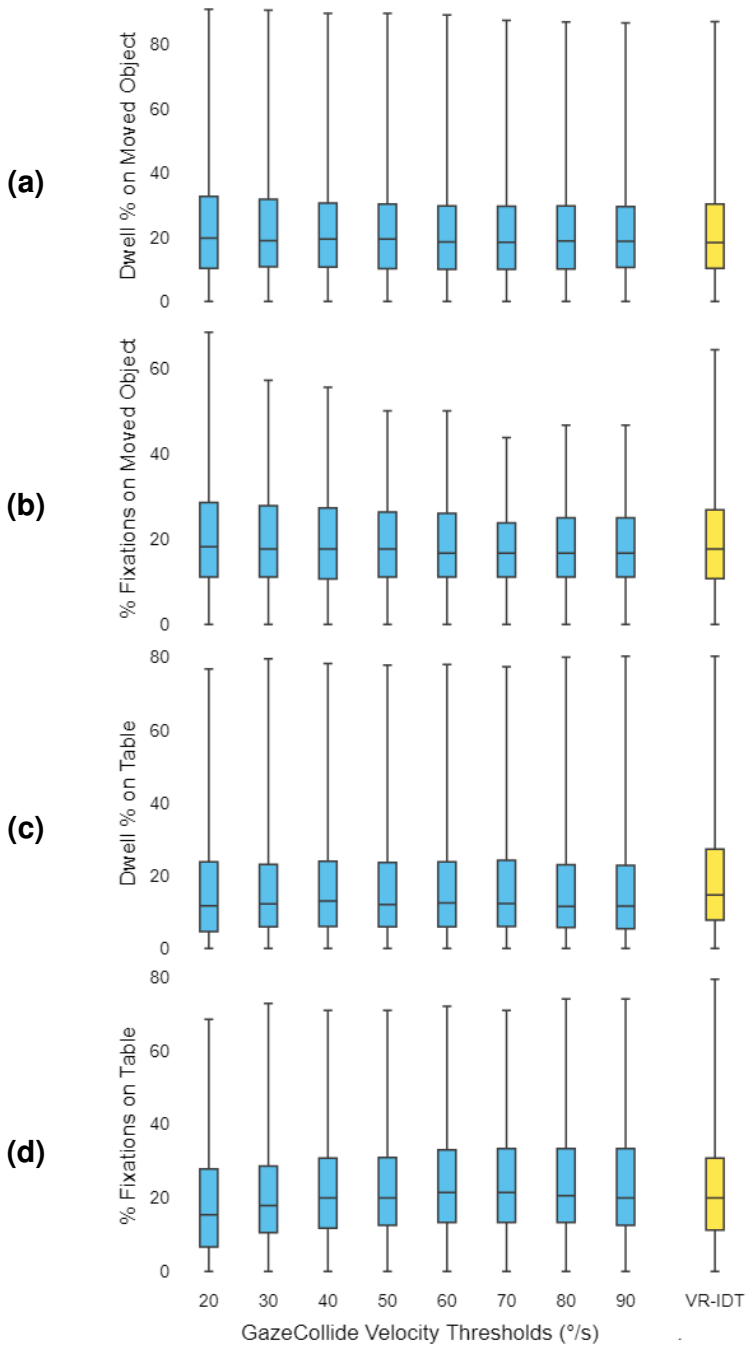


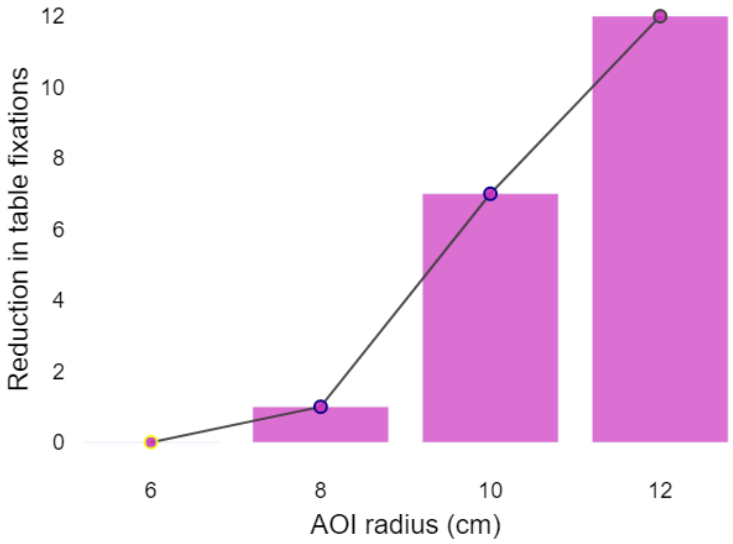
Figure 3.5. Box-and-whisker plots of GazeCollide velocity thresholds and VR-IDT across different eye movement measures: (a) the dwell proportion on the moved object, (b) the percentage of fixations on the moved object, (c) the dwell proportion on the table, and (d) the percentage of fixations on the table. No significant differences were found between boxes for each measure.

3.3.3 Size of AOI spheres

After filtering out data that did not record properly, the number of valid Object Calibration targets for determining the size of AOI spheres was 40 from 8 separate participants. Figure 3.6 provides two visualisations of how changing AOI size and location affected the number of erroneous table fixations from all participants' fixations pooled. Figure 3.6a shows that there was a sharp reduction in the number of table fixations when increasing the size of AOI sphere from 8 cm to 10 cm. Figure 3.6b shows that the majority of table fixations were in locations on the table further away from the viewer.

Visualisation of gaze data (examples in Figure 3.7) confirmed that participants were looking towards configuration objects for the data shown in Figure 3.6. Visual inspection of AOI sizes also suggested that 10 cm and 12 cm radii may be too large for locations closer to the viewer, blocking gaze points intended for the table beyond the object.

(a)



(b)

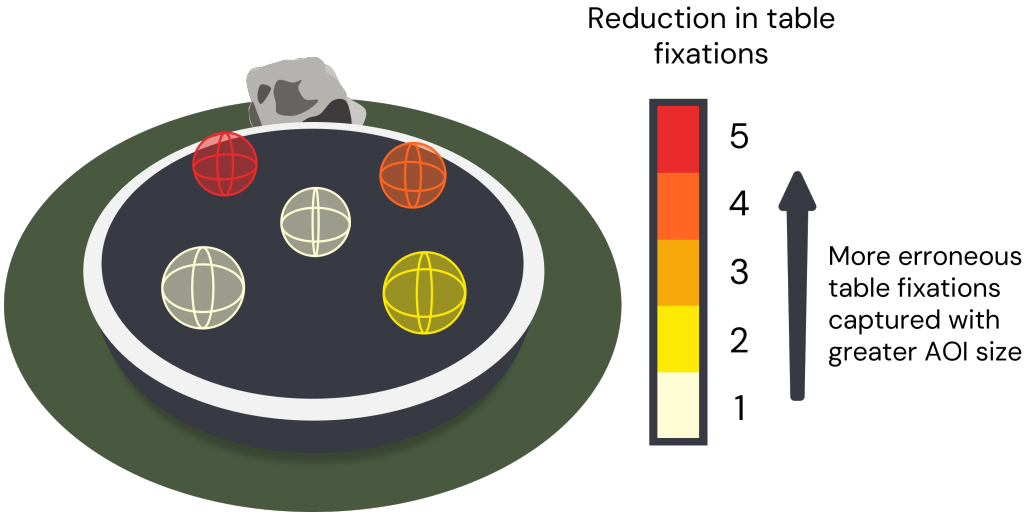


Figure 3.6. Visualisations of the reduction in erroneous table fixations when changing the radius of the AOIs around configuration objects. Note that reduction in table fixations was based on all fixations from all participants. (a) Cumulative reduction in table fixations by AOI radius, such that 8 cm refers to an increase from 6 cm to 8 cm, showing a reduction of one table fixation. (b) A three-dimensional visualisation, from the egocentric perspective shown in Figure 2.1B, of the number of table fixations reduced depending on the location on the table. Wire spheres represent AOI spheres at different locations on the circular table. Sphere colour represents the number of table fixations reduced at that location, regardless of AOI size.

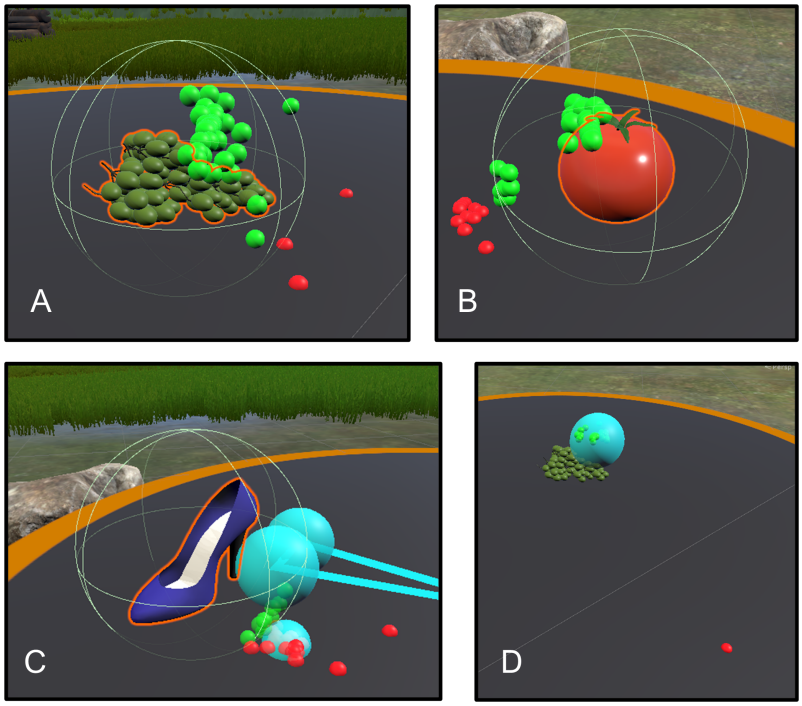


Figure 3.7. Close-up images with visible AOI spheres (green wire spheres), gaze points (red and green dots), fixations (blue spheres) and saccades (blue lines between spheres) as visualised in Unity3D. Red gaze points were registered as being on the table; green gaze points as being on a configuration object. AOI spheres are all 8 cm radius. (A) Gaze points scattered on the table around the object. These were not included in a fixation. (B) An example of an ambiguous cluster of gaze points (red dots) that registered as a fixation point (blue fixation sphere omitted for visibility), nearby another fixation that registered on the configuration object. These table points would be captured by increasing the AOI radius to 12cm. (C) An example of a fixation registered to the table (bottom blue sphere with red gaze points) that should have been registered to the object. These gaze points would have collided with the sphere if the AOI radius was increased to 10cm. (D) An example of an erroneous gaze point, likely due to eye tracking error. These were automatically removed by the GazeCollide algorithm.

3.3.4 Acceptability and usability

All participants found the training instructions clear and felt they understood how to perform and navigate the task well following training and practice, including all participants with MCI.

Four participants commented that some objects were difficult to see or distinguish. These objects were replaced, shown in Figure 3.8.

Sixteen participants commented on the difficulty of the task, including all those who underwent the five-object version and all patients with MCI. Two participants specifically mentioned that it felt like there were too many objects to remember in the five-object condition.

3.3.5 Researcher observations

There was one incident of a participant experiencing motion sickness. However, this was short-lived and mild. Otherwise, participants did not have any tolerance issues with the task. Nevertheless, a screening question was added to the protocol to check for pre-existing sensitivity to motion-sickness or vertigo.

The boundary warning system worked appropriately in preventing participants from any safety risks: this was only required for one incident each in two participants. In other participants, the boundary warning system was not required.

No participants showed signs of difficulties interacting with the task by the end of the practice trials. This included difficulty with orientation after teleporting and difficulty with using the controller to point and click. Three older participants failed the practice criterion once, and one failed it twice, all in the five-object group.

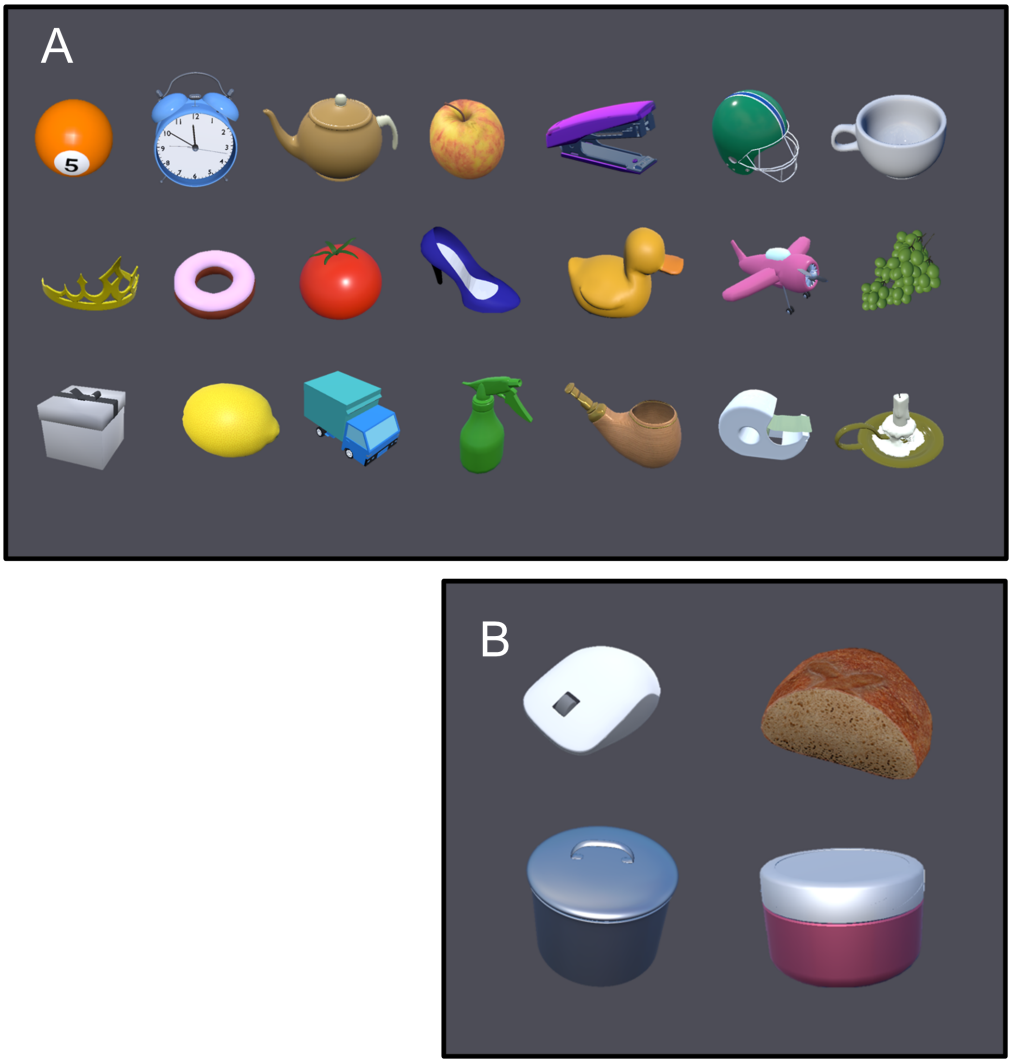


Figure 3.8. Configuration objects. (A) Objects used in the latest version of the task. Objects were selected based on distinguishable appearance including varying colours. Objects were not repeated in successive trials and objects of similar colour or shape did not appear in the same trial (e.g., the 5-ball, apple and tomato never appeared in the same trial). (B) Objects that were replaced or removed following participant feedback.

3.3.6 Memory performance

Participants' percentage of correct trials by four- and five-object groups are shown in Figure 3.9. There was evidence of a reduction in performance for older participants with five configuration objects, compared to older adults with four objects and younger adults with either number of objects ($group\text{-}object\ interaction:\ F(1,\ 19) = 3.07,\ p = .096$; $one\text{-}tailed\ post\ hoc\ \text{t-}test\ of\ four\text{-}\ vs\ five\text{-}objects\ in\ older\ group:\ t(11) = -2.232,\ p = .042$) as shown in Figure 3.9a.

Splitting by condition, older participants did not score significantly greater than chance in the five-object group in any condition except Stay-Rotate (Supplementary Table B.1; $condition\text{-}object\ interaction:\ F(5,\ 45) = 2.22,\ p = .069$). However, older participants did score significantly greater than chance in all conditions in the four-object group (Supplementary Table B.1). These results were supported by visual inspection of Figure 3.9c, which showed particularly low performance between four- and five-object groups in Walk-Rotate and Teleport-Rotate conditions.

No difference was found between four- and five-object younger groups ($F(1,\ 10) = 0.06,\ p = .81$), nor was there a condition-object interaction ($F(5,\ 50) = 1.46,\ p = .22$). This suggests no difference between four- and five-object younger groups in task performance, regardless of condition. These results were supported by visual inspection of 3.9a and 3.9b, which showed a similar performance in four and five object groups across conditions. All four- and five-object younger participants performed significantly greater than chance (Supplementary Table B.2).

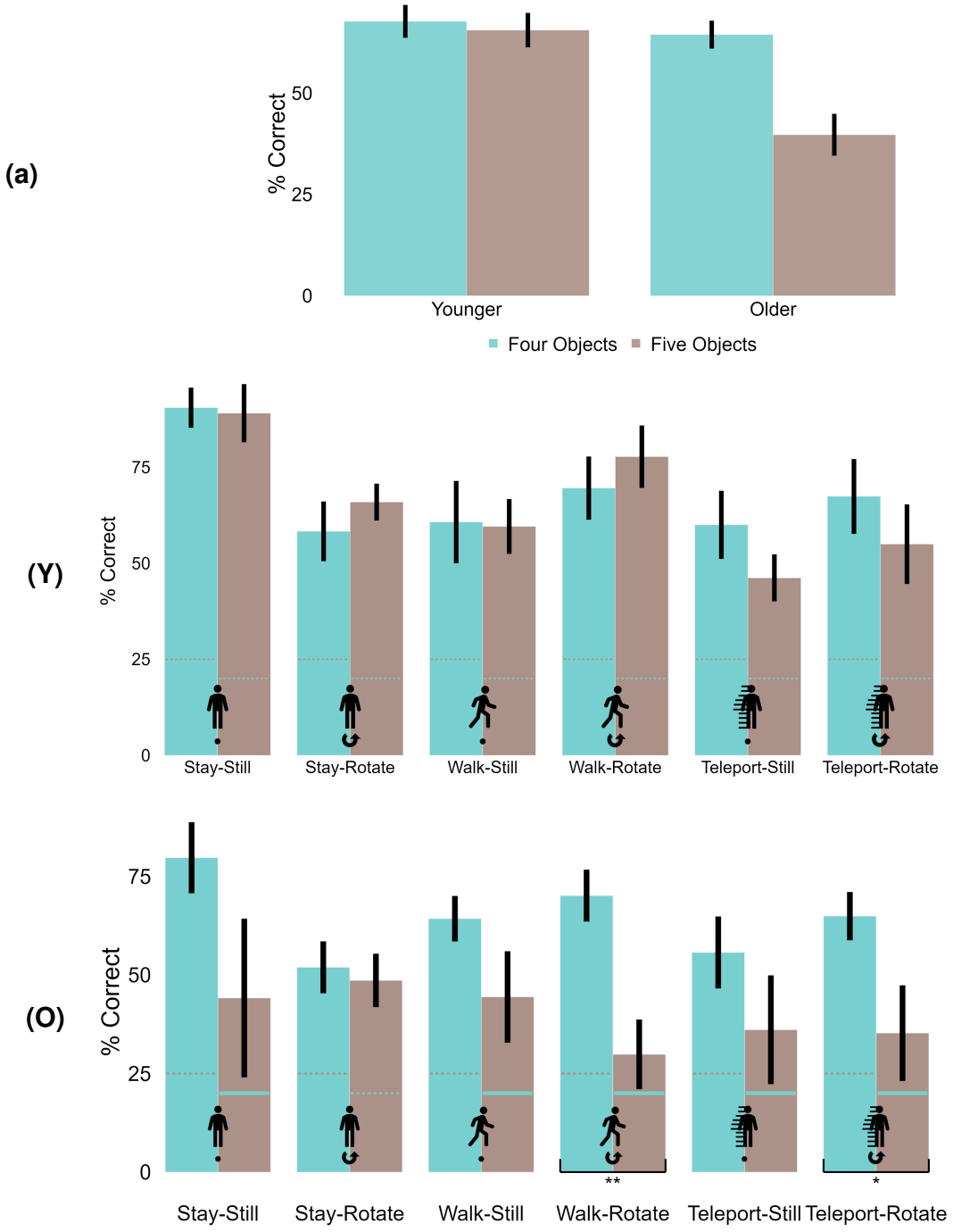


Figure 3.9. Performance by number of configuration objects. (a) All conditions combined, an interaction between group and number of configuration objects. (Y) Younger group split by condition. All scored above chance as denoted by the dotted lines at 25% and 20% for four- and five-object bars, respectively. (O) Older participants split by condition. Solid chance lines represent groups that did *not* score significantly higher than chance, $p > .05$. Significance brackets are results from post hoc independent-samples *t*-tests, * $p < .05$, ** $p < .01$.

3.3.7 Eye movements and memory performance

Figure 3.10 shows the difference between between correct and incorrect trials in the dwell proportion on the moved object at retrieval, split by condition . In both younger and older groups, participants had significantly greater proportion of fixation time ('dwell proportion') per trial on the moved object at the second viewing period for trials that they correctly answered compared to those they incorrectly answered ($older$: $F(1,\ 3) = 151.1,\ p = .001$; $younger$: $F(1,\ 3) = 51.0,\ p = .019$). This suggests that both younger and older adults looked at the moved object more when they were correct compared to incorrect, indicating an association between moved object viewing and task performance.

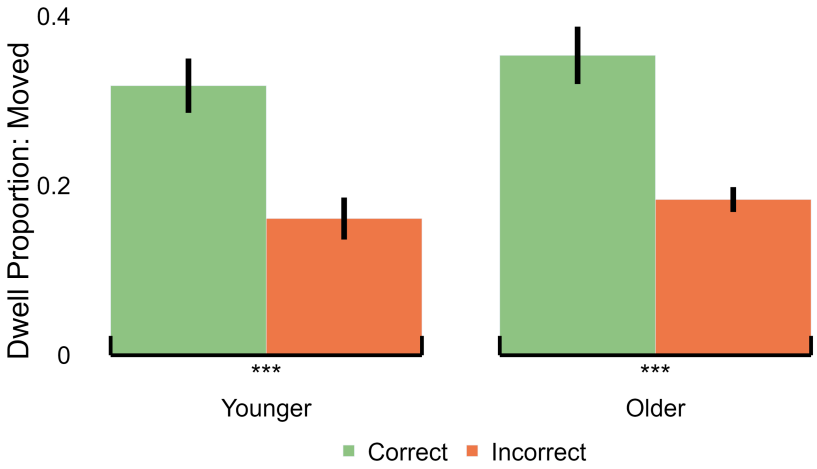


Figure 3.10. Dwell proportion on the moved object at retrieval by group and trial correctness. Brackets at the base of bars represent pairwise significance from hierarchical mixed-effects models. *** $p < .001$.

3.4 Discussion

The study tested the feasibility of a novel spatial memory paradigm in immersive virtual reality (iVR) with integrated eye measurement for further diagnostic research in older and younger adults. This was achieved by assessing measures of technical functioning, data quality, participant experience, memory performance, and eye movements in healthy younger and older adults. Results suggest that the task was appropriate for further testing with some modifications.

Performance feasibility. There was evidence that the five-object version of the task was too difficult for healthy older adults to become a potential test for early Alzheimer's-related memory impairment. These participants scored close to chance in several conditions, and mentioned the difficulty of the task more often than the four-object group. There seemed to be a floor effect for older five-object participants, including in the Walk-Still condition, which needed to allow high enough performance in healthy older adults to show comparatively decreased performance in an early Alzheimer's disease patient group.

Older adults were specifically affected by an increase in the number of configuration objects. This effect on memory performance may have been because fewer object identities and locations need to be encoded and held in working memory, which may boost performance in older adults but not younger (De Beni and Palladino, 2004). Despite improved performance in the four-object group, several older adults still reported finding the task difficult. Other options for making the task easier were not explored, but an additional boost to performance could have come from increasing the viewing time. This was not tested here; the viewing period was already more than double the original paradigm (3 seconds in Simons and Wang, 1998) and increasing this further would either reduce the number of trials, which may have sacrificed statistical power in detecting group differences, or increase the burden on participants. Based on the improved

memory performance in the four-object group, the decision was made to keep viewing time at 7 seconds but continue with the four-object version for further testing, despite the original paradigm involving five objects.

Participants with MCI also found the task difficult, although this was expected and was the desired effect, at least for those with positive biomarkers. This group did not express any particular difficulties in understanding how to perform the task compared to healthy participants, based on qualitative feedback. In fact, no changes were made to the task after beginning testing in patients.

Fixation identification and technical feasibility. Adjustment to the AOIs surrounding objects was necessary given the number of erroneous fixations, particularly for objects further away. However, enlarging AOIs for closer configuration objects may have resulted in false fixations that were intended for the table behind them. The obvious solution was to have larger AOIs for objects further away, compared to closer ones. This was achieved by scaling AOI size between 8 cm and 10 cm depending on the distance of the object from the viewpoint. Changing to this graded approach reduced erroneous table fixations by 7 of 13 compared to using uniform 8 cm (Figure 3.11).

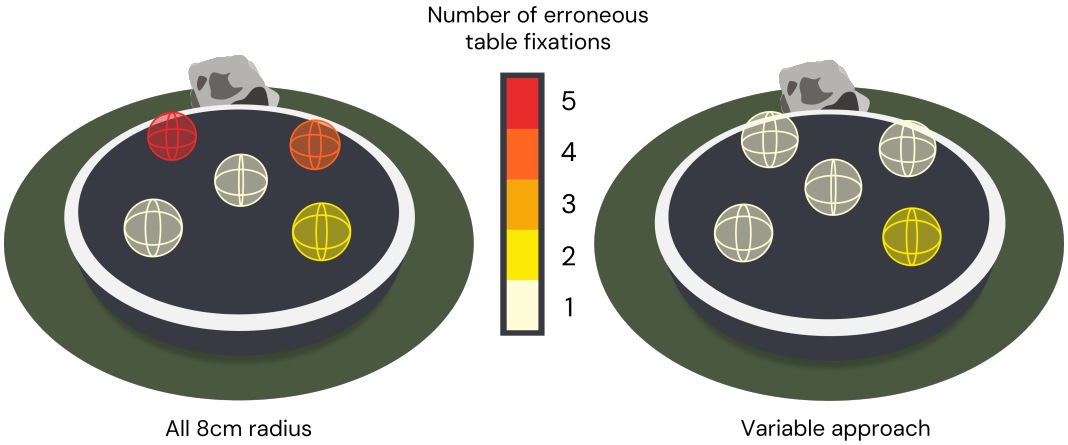


Figure 3.11. Variable adjustment to size of AOI spheres reduced the number of erroneous fixations on the table.

For identifying fixations, GazeCollide labelled eye samples an order of magnitude faster than the VR-IDT algorithm (Llanes-Jurado et al., 2020). This is because GazeCollide avoided using a moving window with nested loops of code, instead taking advantage of vectorized operations on all samples simultaneously. An additional advantage of GazeCollide was its suitability to an object-location task with key AOIs spaced apart. As mentioned, this made fixation samples on the objects central to classification of fixations, rather than a post hoc process. One could see a disadvantage to this approach if attempting to be more precise around intra-AOI fixations. Indeed, a velocity threshold was introduced to ensure saccade samples were not falsely labelled as fixations for saccades within AOIs. However, for the current hypotheses, ensuring that fixation dwell time per AOI was accurate was more important than within-AOI eye movement detection. Of course, this was only as good as the AOIs themselves, hence the calibration of the AOI sizing.

Note that there are several modern eye-event algorithms that show impressive classification performance when trained on expert-labelled data (e.g. Zemblys et al., 2019). The advantage of these, apart from their accuracy, is their ability to generalise across levels of noise and equipment, assuming cer-

tain constraints. However, the trade-off is (a) the complexity of their approach, which may be a black box in the case of artificial neural networks, and (b) their processing speed, development time and/or training time, which are often not reported.

Only one previous fixation algorithm—VR-IDT—was used to calibrate and compare against the new GazeCollide algorithm because so few have been specifically adapted for VR. However, a useful direction in this field would be to adapt the best performing eye-event classifiers to iVR eye tracking data, perhaps by further training pre-trained models. Alternatively, a common way of validating eye-tracking hardware is by using an artificial eye and simulating gaze behaviour, then calculating accuracy and precision based on the protocol of the artificial eye (Wang, Mulvey, et al., 2017). Previous independent assessments of VR eye-tracking systems using artificial eyes have been published (Lohr et al., 2019) including with the HTC Vive Pro Eye (Sipatchin et al., 2021), but for different purposes. Development of an artificial eye to validate GazeCollide or other fixation algorithms was beyond the scope of this research, but could be a valuable addition to the field.

Eye movement measures. Results suggest that examining eye movements alongside task performance may be useful for detecting differences between groups. Indeed, there was a difference in moved object viewing between correct and incorrect trials in both older and younger adult groups. This is consistent with preferential viewing effects found in previous literature (e.g. Yeung et al., 2019). The advantage of eye tracking measures is that even with such few participants, enough statistical power was achieved to detect differences. Caution must still be taken in interpreting these results, as the data were generated from a small number of participants and therefore any conclusions will be over-fit to the sample. Still, we can observe differences within groups and across trial correctness, supporting the use of these eye movement measures as potential

tests for between-group differences.

Given the limited precision of the eye-tracking, eye movement metrics that rely on spatially- or temporally-precise measurements were unsuitable for study. For example, microsaccadic measures such as dynamic overshoot (Bahill et al., 1975), which have been shown to be abnormal in AD (Kapoula et al., 2014), are usually defined as tiny eye shifts $<1°$, whereas the precision of this eye-tracker was around 1°. In fact, most saccade measures were unsuitable, including prosaccade latency, maximum saccade velocity, and saccadic intrusions. Accordingly, measures of saccades were avoided for the main results of the larger study described in the following chapters, focusing instead on fixation AOIs and sequences.

Limitations. Existing literature offers limited guidance on robust feasibility study design, with available options predominantly focusing on intervention studies (Downs and Black, 1998). Such a rarity of standardised methods is understandable, given that feasibility testing varies greatly based on study details. Indeed, the majority of this chapter necessarily focused on customised approaches to assessing feasibility. Naturally, these novel methods have some limitations, which are discussed here.

Firstly, in comparing four- and five-object versions of the task, a within-subjects design may have been more suitable. A counterbalanced repeated-measures study would have provided more statistical power and would have reduced noise produced by individual differences. This design was avoided because a comparison between number of objects was not originally planned: the task was initially designed with five objects, consistent with earlier paradigms (Burgess et al., 2004; Simons and Wang, 1998) and without any evidence that fewer objects increased task performance. After older adults in the five-object task were found to be performing close to chance, a four-object version was tested for comparison. At this point, a within-subjects design with the same

participants may have been biased by practice effects, so a between-subjects approach was taken. Despite this, enough evidence was found with a between-subjects design to choose the four-object version of the task.

I have discussed the importance of calibrating task difficulty away from floor effects, but this issue could have been avoided by collecting a continuous measure per trial instead of modelling performance as a binomial distribution. For instance, previous studies have measured the absolute error from participants replacing objects to their original position (Mou and McNamara, 2002; Postma and De Haan, 1996). This could have appended or even replaced the object-selection phase, providing a potentially powerful measure of performance. However, this option would have been a departure from the original paradigm, and may have increased testing time, further increasing burden on participants. Moreover, continuing with a Bernoulli distribution per trial has the advantage of examining eye movements in correct trials against incorrect ones, as we have seen in Figure 3.10.

Although an attempt was made to systematically calibrate task features, the method for determining the size of AOI spheres was not perfect. It was difficult to be confident that gaze points or fixations on the table nearby were actually intended for the object, even with the Object Calibration design and subsequent visual inspection. Indeed, some participants may have actually fixated near configuration objects instead of directly on them (Rösler et al., 2005). Increasing the size of AOI spheres to account for this could have obscured viewing behaviour that involved gaze points nearby or in the spaces between objects, which varied between groups as described in subsequent chapters. An additional instruction could have been added to Object Calibration to look *directly* at the objects as they appeared, in an attempt to reduce the uncertainty around table fixations. However, the instructions were deliberately chosen without this to allow for more natural viewing of each object, improving comparability to the

main task. Regardless, the task was inherently limited in its ability to represent gaze behaviour around the peripheries of objects while it involved hard AOI thresholds, although this type of analysis may be precluded by the quality of the eye-tracker anyway. Still, if I can demonstrate some diagnostic value of using eye-tracking from this imprecise setup, then it would support the use of lower-quality eye-tracking systems for clinical use, being cheaper and potentially more widely available than alternatives.

Conclusion. In sum, the task and its host system had room for greater precision in the broad sense, but evidence supported the paradigm as being ready and suitable for further testing. The breadth of the task in detecting subtle eye movement changes was limited due to the precision of the eye-tracker. However, the combination of spatial memory testing with broad fixation measurement was still a novel direction for Alzheimer's diagnosis research.

Chapter 4

Differences in spatial memory, confidence and eye movements in healthy ageing and mild cognitive impairment

Abstract

The main results of the thesis are presented, organised by modality (task performance, confidence, and eye movements).

The dataset. The data collection procedure is first described. Participants comprised four different groups: (1) healthy younger adults, (2) healthy older adults, (3) people with MCI but no signs of Alzheimer's biomarkers, (4) and people with both MCI and presence of AD biomarkers. A description of the dataset is provided, including missing or removed data and any differences between groups in demographics and neuropsychological test results.

Hypotheses. I hypothesised that the Walk-Still condition would show the greatest differences between MCI biomarker groups for both task performance and eye fixation measures, with a more condition-invariant impairment in memory monitoring abilities for biomarker-positive participants. Ageing differences were not expected to show modulation by spatial factors across different conditions, but greater memory effects were expected in younger participants than their older counterparts.

Summary of results. The Walk-Still condition was *not* the most discriminating condition, and neither was task performance more broadly. Rather, it was

eye movements in the Stay-Still condition that showed the greatest differences between MCI biomarker groups, particularly for the proportion of time spent fixating on the stationary objects and the table during the retrieval period. Memory monitoring abilities may have been reduced in biomarker-positive patients compared to healthy older adults, but not compared to biomarker-negative MCI participants. In healthy age groups, younger adults scored higher on the task across all conditions, as expected. These younger participants looked at the table more during encoding and retrieval, with the latter effect *negatively* associated with task performance, contrary to expectations.

4.1 Introduction

This is the first of four chapters describing and exploring the dataset from the main study in the thesis. Together these chapters can be considered an extended results section, starting with descriptions of the participant sample and the dataset collected from them, before detailing a series of analyses with the predominant focus of identifying predictors of participants with both Mild Cognitive Impairment and positive biomarkers for Alzheimer's disease (MCI+). Starting in this chapter, MCI+ participants were compared to participants with MCI and negative biomarker results (MCI-) across spatial memory, confidence and eye movement measures. Additionally, healthy ageing effects were investigated by examining differences in task measures between healthy younger and healthy older adults, with this latter group doubling as another control group to compare against MCI+ participants. A description of the data collection and analysis procedures are provided in the next section, including detailed hypotheses.

This first results chapter tests the main hypotheses of the study—with some related post hoc analyses—by comparing task measures between groups and conditions, and examining associations between eye movement measures and spatial memory performance. All eye movement measures in this section were viewing-level or trial-level summary statistics, such as the proportion of fixation time on the moved object during the retrieval period.

An alternative method of eye movement analysis will be introduced and examined in Chapter 5. This involved modelling the time-dynamics of viewing patterns over an averaged viewing period per participant, per condition. Two alternative statistical methodologies were used to test for group-level differences in how viewing behaviour changed within viewing phases.

Finally, Chapter 6 examines the diagnostic value of various classification models derived from task measures as compared to traditional neuropsychological tests. This analysis directly assesses how the task compares to stan-

dardised cognitive tests in its ability to identify participants at the highest risk of AD.

Additionally, Appendix A introduces a new potential method for detecting encoding differences between groups based on clustered scan-paths on similar object configurations. This technique was not appropriate for valid study results, but provides a promising methodology for future analyses or experiments.

The current chapter is the most results-dense section of the thesis, containing numerous comparisons across several sub-sections. Accordingly, frequent summaries of notable findings will be provided immediately prior to more complete descriptions of results. A narrative summary of all findings in the chapter can be found in the final section of the chapter.

4.2 Data collection, participants and the dataset

4.2.1 Participants

A convenience sample of four groups of participants were recruited to the study. All participants were subject to the following exclusion criteria:

- A visual impairment that could not be corrected-to-normal, including colour visual deficiency (colour-blindness);
- Significant mobility problems or injuries that may cause problems with the task (such as difficulty balancing while wearing the head-mounted display);
- Significant mental health or psychiatric disorders that may affect task performance;
- Difficulty with verbal or written instructions due to communication needs or language barriers;
- Any other diagnosed neurological conditions including cerebrovascular disease.

Healthy younger adults ('Younger') were recruited through University College London's (UCL) Division of Psychology and Language Sciences (PALS) Subject Pool by advertising the study through an online website, where volunteers could self-select whether to participate. In addition to the above exclusion criteria, younger participants were only included if aged under 30 years old.

Healthy older adults ('Older') were recruited via Join Dementia Research (joindementiaresearch.nihr.ac.uk), through social media advertising, and through University Hospitals Sussex NHS Foundation Trust (UHSx). The minimum age for this group was determined by the youngest participant in the patient groups, described next.

Patients with a diagnosis of Mild Cognitive Impairment (MCI) were recruited from UHSx following initial assessment for MCI by a Consultant Neurologist according to the Peterson criteria (Petersen, 2004). All MCI participants presented

evidence of memory impairment as assessed by the Addenbrooke's Cognitive Examiniation (ACE-III), and hence were diagnosed with the amnestic MCI subtype. MCI participants were recruited if they had undergone or were awaiting clinical biomarker testing for signs of early AD according to IWG2 criteria (Albert et al., 2011). All but one of these participants received this testing via lumbar puncture and assessment of cerebrospinal fluid (CSF) for signs of low amlyoid-β_{42} (assessed by the ratio of $A\beta_{42}$ to $A\beta_{40}$), high total tau, and high phosphorlyated tau-182 (p-Tau). One participant underwent Positron Emission Tomography (PET) scanning with a radioactive tracer for Aβ and tau depositions. Clinically standardised thresholds for biomarkers were used to categorise MCI patients into biomarker-positive or -negative (NICE, 2018).

The optimal sample size was calculated as 28 per healthy group based on standard errors from previous versions of the paradigm (Simons and Wang, 1998) and effect sizes from a similar study between healthy older and younger adults with walk and teleport conditions (Castegnaro, 2021). For calculating the minimum sample size for MCI groups, the benchmark has arguably been set by the path integration iVR test (Howett et al., 2019), which separated biomarker groups with a very large effect size (Cohen's $d \approx 2.4$). This effect size requires 5 participants per group to detect a significant difference at an alpha level of 0.05. However, MCI participants were recruited as far beyond this number as possible in the available time.

Note that these sample sizes were based on power calculations for task performance as measured by proportion of correct trials. However, confidence rating and eye movement measures had greater statistical power for detecting group differences due to hierarchical modelling of trial-level data (see Section 4.3.2). Power calculations were performed using G*Power (Faul et al., 2007).

The study was undertaken in line with regulations outlined in the Decla-

ration of Helsinki (WMA, 2013) and was approved by the London-Bloomsbury Research Ethics Committee (21/LO/0561) and Health Research Authority. Participants recruited under these ethics were offered £20 towards travel reimbursement. Recruitment of healthy adult participants was approved under the UCL Psychology and Language Sciences Divisional Research Ethics Committee (SHaPS-2018-JK-027). These participants were paid either £7.50 or £11 per hour, depending on the year, in accordance with departmental policy at the Institute of Cognitive Neuroscience.

4.2.2 Neuropsychological examination

Participants completed a battery of neuropsychological tests in addition to the iVR task. The tests were chosen based on their widespread clinical use for assessment of MCI and AD, and provided clinical best-practice comparisons against iVR metrics for discriminating MCI+ from control participants in Chapter 6.

Participants were asked to complete data collection across two separate sessions: the first involved the iVR task, and the second the neuropsychological battery.

Premorbid functioning. Not all neuropsychological tests were included as direct competitors with the iVR task for MCI+ discrimination. For example, the National Adult Reading Test (NART; Nelson and Willison, 1991) was included as a test for estimating premorbid cognitive functioning. This is a key component of neuropsychological assessment because general intelligence quotient (IQ) scores have been shown to affect results of other cognitive tests (Ratcliff et al., 2011). Significant group differences in IQ could therefore confound diagnostic results. However, controlling for IQ is not possible with MCI patients because cognitive impairment affects IQ scores (Starr and Lonie, 2008). Instead, an estimate of pre-impairment, or *premorbid*, IQ must be estimated indirectly. Reading ability for irregular words has been found to significantly correlate with

IQ in healthy adults and be relatively preserved in patients with dementia (Nelson and McKenna, 1975). The NART was included for MCI participants and healthy older volunteers, with the aim of sampling groups that are not significantly different in their NART scores. The NART was not included for healthy younger participants because (a) as a test of premorbid functioning, it is not relevant to the healthy younger group, and (b) it is not valid for testing non-native English speaking populations (Nelson and Willison, 1991), which accounted for the majority of the Younger group (Supplementary Figure B.2).

Addenbrooke's Cognitive Examination. The Addenbrooke's Cognitive Examination (ACE-III; Mathuranath et al., 2000) was included when testing healthy older adults to allow comparison with MCI volunteers who received the same testing during clinical assessment. The ACE-III screens for several areas of cognitive dysfunction, providing an overall composite score based on sub-scores of attention, verbal fluency, language, perception, visuospatial functions and memory. The ACE-III is used for screening early signs of dementia, as validated against a healthy older normative sample. Younger performance on ACE-III was not reported because ACE-III testing was only relevant to MCI group comparisons.

Tests of delayed recall. All remaining neuropsychological tests were included due to their evidence for discrimination of people with AD from healthy control participants, some of which are used in clinical neuropsychological assessment of AD.

Two tests of delayed recall were included. The first was from the Free and Cued Selective Reminding Test (FCSRT; Buschke, 1984). This required participants to repeatedly learn sixteen images alongside cued associations. This test provided a verbal recall comparator with an object location component: participants were presented with the images four at a time, and could encode the spatial location of the image (top-left to bottom-right) as well as the name and object category. The delayed free recall measure, which has previously

shown a degree of diagnostic accuracy for AD (Grande et al., 2018) required participants to freely recall as many of the sixteen images as possible after a 30 minute delay, without verbal cues.

The second test of delayed recall was from the Rey-Osterrieth Complex Figure Test ('Rey'; Corwin and Bylsma, 1993). Subjects were asked to copy a detailed figure without any warning to memorise the drawing. They were then immediately asked to redraw the figure from memory. And finally, after 25 minutes, they were asked to draw the figure from memory for again. This final delayed recall measure has been used in memory impairment assessment for many decades and has some evidence for its discriminability of AD from healthy age-matched controls (Bigler et al., 1989).

Tests of processing speed and executive function. Two tests of processing speed were included: the Trail-Making Test part B (TMT B; Reitan, 1958) and the Digit Symbol Substitution Test (DST; Wechsler, 2019). The TMT B is an attention-switching task that requires executive functions to plan and execute a timed line drawing test. Participants must draw a continuous line between an irregular arrangement of circles containing numbers and alphabetical letters. This is commonly used as an assessment for AD, with higher times indicating greater likelihood of cognitive impairment (Lafleche and Albert, 1995). The DST requires participants to place as many corresponding symbols beneath a sequence of single digit numbers. The task is straightforward but time-consuming; subjects must complete as much as possible within 2 minutes. Participants with AD have performed lower than healthy controls and participants with subjective cognitive decline on the DST (Tsatali et al., 2021).

The 4 Mountains Test. Finally, the 4 Mountains Test (4MT; Hartley et al., 2007) was included as a test of allocentric spatial memory. The test requires participants to memorise a three-dimensional scene of four mountains on a two-dimensional screen. The mountains vary in size, shape, and spatial location. After 5 seconds, participants are presented with four scenes of four mountains,

one of which is the same mountain configuration they studied but from a shifted viewpoint. The goal is to match the learned mountain arrangement to the correct option; the other three foil options have different spatial arrangements of the mountains. Cognitively, this task is most directly comparable to the iVR task in this study as it is designed to test allocentric spatial memory. The 4MT has found promising results for discriminating healthy older adults, participants with MCI and patients with AD (Chan et al., 2016).

Neuropsychological test scores were compared between groups using Analyses of Variances (ANOVAs) for main effect of group on test metric, and post hoc *t*-tests for pairwise comparisons, with a chi-squared test to compare balance of males and females between groups.

4.2.3 Exclusions and missing data

Three healthy older adults and three MCI+ participants were excluded because they could not perform the iVR task without prompts by the end of the practice block. A further two MCI+ participants were excluded after consenting due to clear difficulties balancing while wearing the iVR headset. These participants did not undergo neuropsychological testing.

Ten participants had missing eye-tracking data due to technical faults. One Younger participant had no eye-tracking data from any trials. This participant was removed from all analyses. One participant from Younger, Older and MCI+ groups, and two participants from the MCI- group, had no eye-tracking data from one full data collection block, also due to technical faults. These participants still had full task performance and confidence data and thus were kept in the dataset, albeit with 6 instead of 9 trials to calculate eye-tracking metrics from per condition. Note that this was accounted for in statistical analysis via hierarchical mixed-effects modelling (see Section 4.3.2).

Of the viewings (encoding and retrieval) that were successfully recorded, 0.35% were removed for having $>$25% missing eye-tracking data. Figure 4.1 visualises the distribution of missingness by viewings.

Some participants did not complete the full neuropsychological battery for a variety of reasons. For example, the Rey test was skipped for one Younger and one Older participant who had both encountered it before; the test required subjects to be unaware of the recall components to be valid. One further participant did not complete the Rey test because of researcher fault. One MCI- participant did not complete the Trail Making Test because they had encountered it recently and may have had a practice effect.

Occasionally, time-constraints led to missing data. For example, four Older participants did not complete the ACE-III for this reason; one of these also skipped the Rey test due to previous exposure.

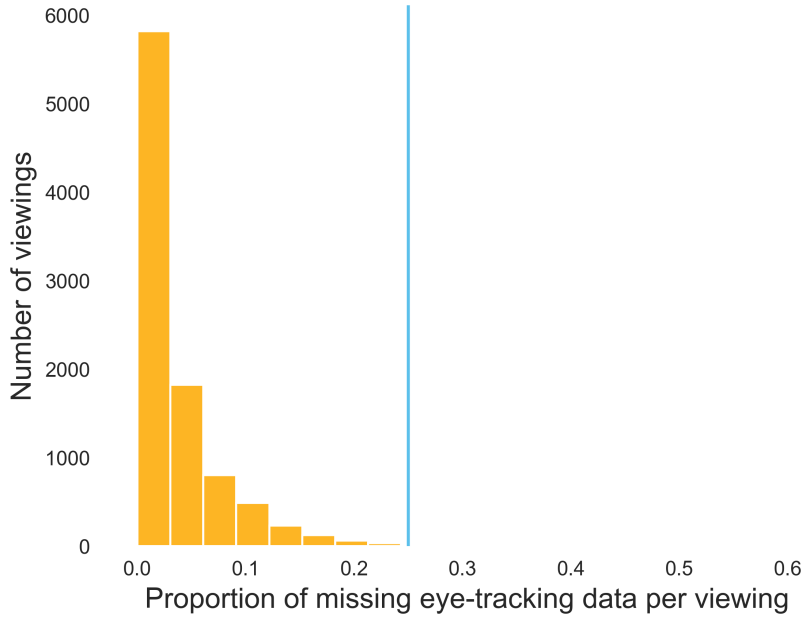


Figure 4.1. Histogram of missing eye-tracking data per viewing. The blue vertical line is the 25% threshold for exclusion of viewings from further analysis.

Nine healthy participants with only one or two neuropsychological tests missing had these scores imputed using multivariate regression imputation. This form of imputation was chosen because only healthy adults needed imputing and significant correlations between neuropsychological test scores have been previously reported in healthy adults.

For each neuropsychological measure with missing data (ACE-III, TMT B and Rey), significant predictors of that measure in healthy participants (Younger and Older) were used to predict missing values. The set of significant predictors per measure was determined by (a) building a multiple regression model with each combination of predictors, (b) randomly removing known values of the target measure to be imputed, and (c) comparing the root mean square error of the predicted and true values to get the set of predictors with the least error. For example, TMT B scores were best predicted by scores on the DST.

The effect of imputation compared to missing data removal was compared on a per measure basis. Imputation of missing data did not change significance

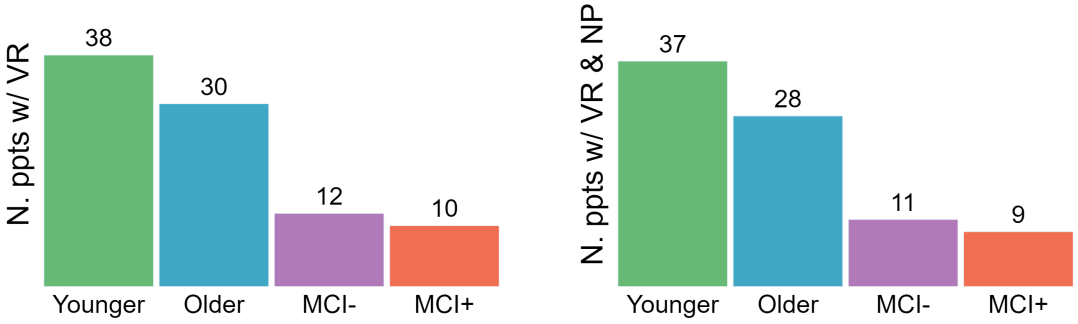


Figure 4.2. Sample size per group. The left plot shows the total participants with VR data, the right plot shows total participants with neuropsychological data as well.

of group comparisons. Note that a full multiple imputation of missing values was not performed due to time constraints.

An additional one MCI+ and two Older participants had no neuropsychological data at all. The two Older volunteers withdrew from the study before the second session for unforeseen and completely random reasons. The MCI+ participant could not be tested for the second session due to a study-wide disruption in data collection unrelated to participant characteristics. Although this missingness disproportionately affected Older and MCI+ groups, data were considered to be missing completely at random. However, too many measures were missing to impute data for these participants. Accordingly, these participants were included in group comparisons of VR task measures, but removed when comparing group neuropsychological results and when estimating classification accuracy in Chapter 6.

The final sample sizes per group are shown in Figure 4.2 for VR data alone and for participants with both VR and neuropsychological data.

4.2.4 Descriptive statistics and neuropsychological test results

Figure 4.3 visualises group similarities and differences in descriptive statistics and neuropsychological test results.

No significant differences in age or years of education were found between Older, MCI- and MCI+ groups $(p > .05)$.

The ratio of male to female participants did not significantly deviate from an equal split $(\chi^2(3) = 6.28,\ p = 0.10)$. However, some caution should be taken when interpreting the data because the groups were not similarly imbalanced by sex. The top-right sub-figure in Figure 4.3 visualises the deviation from 50% males: the smaller the bars, the more balanced the sexes. We can see that MCI groups had more men than women, particularly the MCI- group, whereas the healthy younger and older groups were the opposite.

No significant differences in scores on the NART were observed between Older, MCI- and MCI+ participants $(F(2,\ 45) = 2.1,\ p = .13)$, suggesting that patient groups were matched to healthy older adults on premorbid cognitive function. A main effect of group on total ACE-III scores was found $(F(2,\ 46) = 31.8,\ p < .001)$. Figure 4.3 shows a significant difference between the Older group and both MCI groups. A small significant difference between MCI groups was also found on ACE-III scores.

A significant main effect of group was found on FCSRT Delayed Recall $(F(3,\ 81) = 24.0, p < .001)$, with post-hoc *t*-tests revealing that healthy younger and older participants remembered significantly more images than both MCI groups.

There was a significant main effect of group on Rey Delayed Recall $(F(3,\ 81) = 20.8,\ p < .001)$. I found that Younger participants remembered significantly more of the Rey complex figure than all other groups, and MCI+ participants scored lower than all other groups. Indeed, Figure 4.3 demonstrates that the pattern of group means is consistent with an Alzheimer's diagnostic test.

A similar pattern of results was observed for the Immediate Recall measure on this test (Supplementary Figure B.1).

In this population, healthy younger participants were significantly faster than all other groups on the TMT B, and healthy older participants were significantly faster than both MCI groups, with no significant difference between MCI- and MCI+ participants. The results for the DST were similar but inverted, although MCI+ participants were not significantly slower than Older participants.

A significant effect of group on 4MT scores was found $(F(3,\ 79) = 7.5, p < .001)$, although post-hoc tests found no differences between Older and MCI participants, with the only post hoc pairwise effects between Younger participants and all other groups.

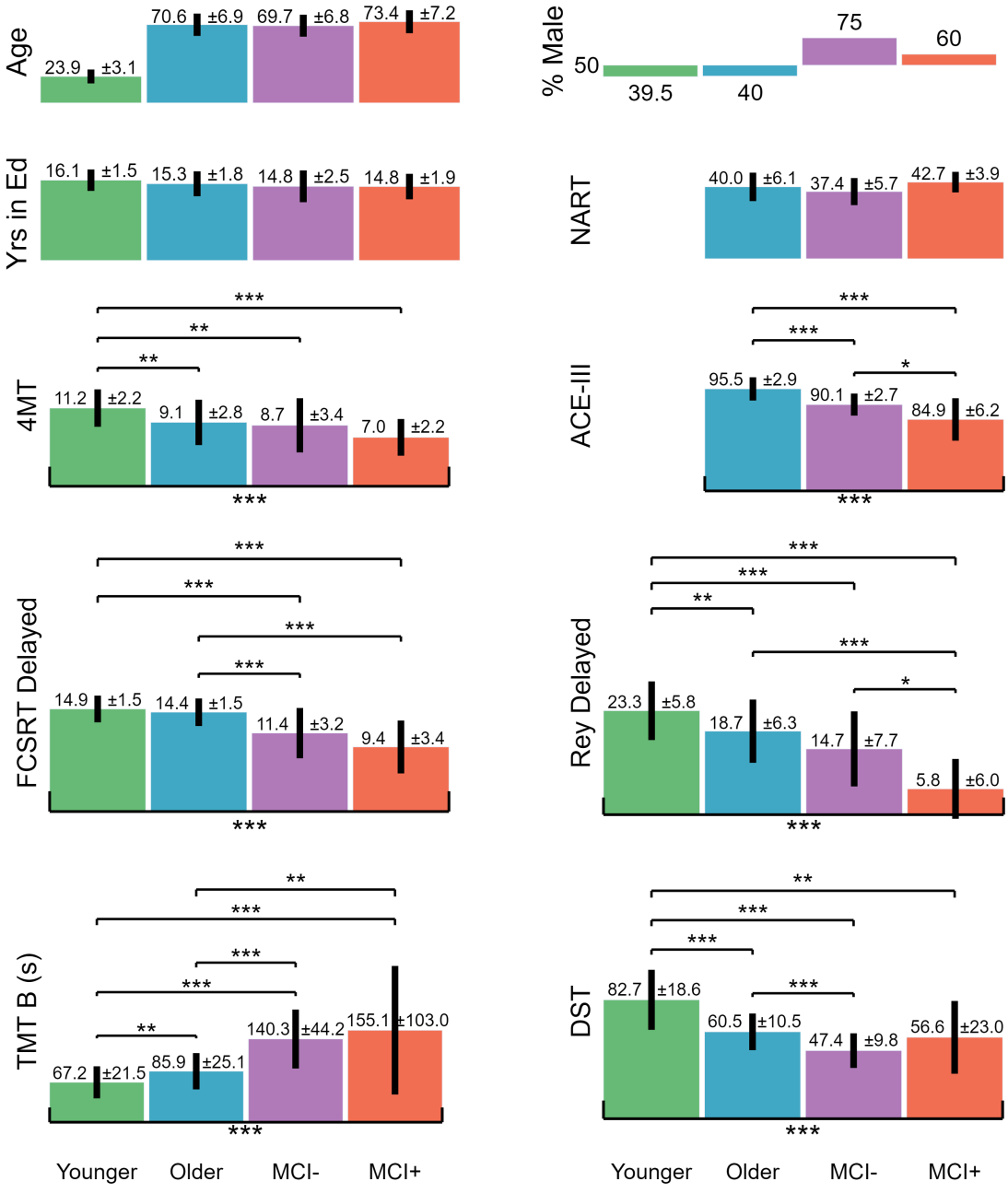


Figure 4.3. Group descriptive statistics and neuropsychological test results. All units are scores, except Trail Making Test B in seconds. Brackets at the base of bars are significant main effects, brackets above bars are post hoc pairwise comparisons. Significant differences in age between Younger participants and other groups are not shown. * $p < .05$; ** $p < .01$; *** $p < .001$; Error bars are standard deviations; see main text for abbreviations.

4.3 Analyses

This section focuses on the differences between participant groups on trial-level outcome measures, with analysis of eye movement associations with task performance. Memory and confidence findings will be described first, followed by several results from eye-tracking measures.

4.3.1 Hypotheses of task results

Before describing the results of this section, hypotheses from the introductory chapter are reproduced and expanded upon.

AD-related hypotheses for memory performance. I predicted that participants with MCI+ would have no benefit to memory from self-motion (walking), or from allocentric spatial consistency with the object positions. This was primarily tested by examining group differences within key conditions, aligning hypotheses with the wider aim of finding the most discriminating conditions for early AD. However, an alternative approach based on factorial comparisons of spatial consistencies is described in Section 4.3.2.1. The consistency of each condition with different spatial factors, and the expected pattern of group differences by condition, is visualised in Figure 4.4. The reader is encouraged to refer to this figure when reading the remainder of this section.

The Walk-Still condition was expected to provide the most discrimination between MCI+ and other participants. This is because there was no availability of an egocentric visual 'snapshot' of the configuration at encoding (because the viewpoint had shifted), but the configuration remained consistent with (a) a self-motion-updated representation of the encoded array and (b) environmental cues (Burgess et al., 2004). Damage to the medial temporal lobe in MCI+ participants was hypothesised to impair these spatial functions, which should boost performance in non-AD groups.

Younger, Older and MCI- participants were expected to perform worse in

the Teleport-Still condition compared to the Walk-Still condition because of the disruption to self-motion. However, MCI+ participants were also hypothesised to perform significantly worse than other groups in this condition because of impairment to environmental (allocentric) representations, which remained consistent in these trials and would have therefore provided a performance boost to healthy participants. MCI+ participants were expected to perform as poorly in Walk-Still and Teleport-Still conditions, unlike other groups.

To test whether these hypothesised impairments were indeed due to problems with self-motion and allocentric processing, table rotation conditions were included. This varied the availability of egocentric strategies, while removing the consistency of the object positions with self-motion and environmental cues. For example, in the Stay-Rotate condition, the object configuration is inconsistent with: the encoded view; self-motion updating; and external environmental stimuli, simply by rotating the table away from the participant. This condition is expected to elicit the lowest performance across healthy participants, consistent with previous results (Simons and Wang, 1998; Burgess et al., 2004). MCI+ participants should perform no worse than MCI- volunteers in this condition if their impairment is due to self-motion updating and allocentric processing.

The Walk-Rotate and Teleport-Rotate trials were hypothesised to show no differences between MCI+ and other groups either, because both conditions were only consistent with an egocentric representation of the encoded object positions; healthy participants would have no performance boost from either self-motion updating or environmental consistency because the table rotated in relation to these. Moreover, these conditions were hypothesised to elicit equal performance across all groups; the alternative would suggest that there is an effect of teleporting that affects performance per se.

Finally, the no-change Stay-Still condition was hypothesised to show no significant difference between Older, MCI- and MCI+ groups due to the availability of same-viewpoint egocentric representations, similar to Move-Rotate

trials.

Healthy ageing hypotheses. The above effects were expected to show between Older, MCI- and MCI+ groups. In addition, a condition-invariant effect of healthy ageing was hypothesised, whereby Younger participants would perform better than Older participants across conditions, with no condition-age interaction. The reasoning for this can be found in Section 1.6 towards the end of the introductory chapter.

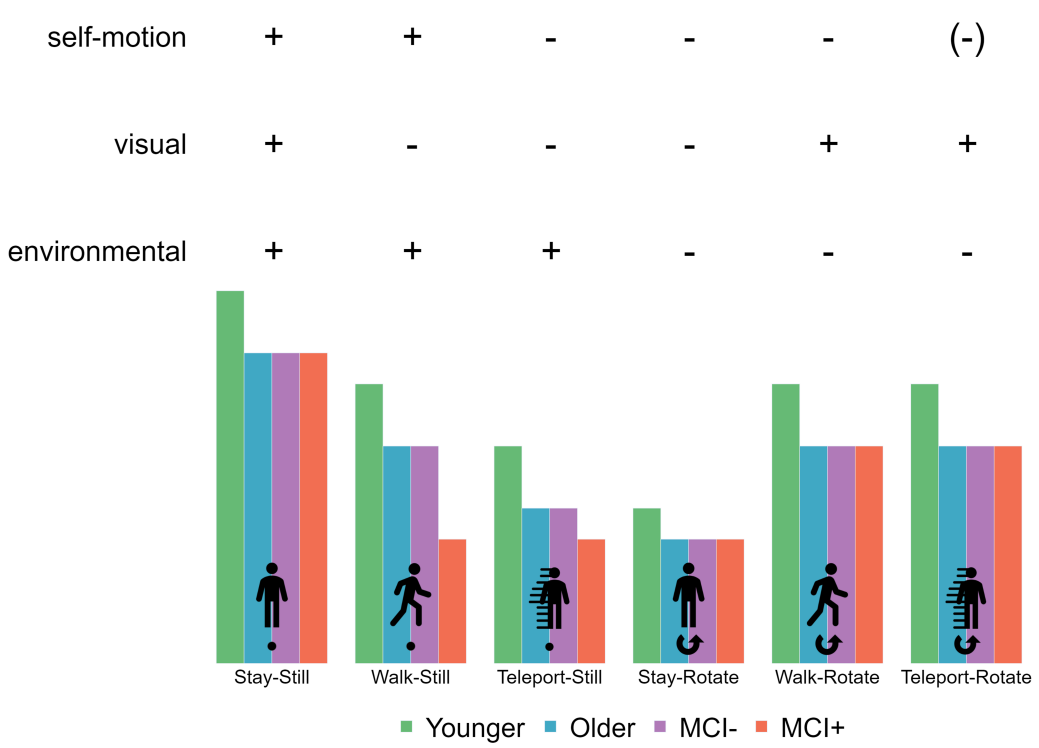


Figure 4.4. Hypotheses by condition and group. The top plot shows consistency of each condition by different spatial advantages. The bottom bar chart shows hypothesised relative scores for memory-related measures (i.e. task performance and memory-related eye movement measures).

Eye movement hypotheses for the retrieval phase. Eye movement measures linked to memory were expected to follow the same or inverse pattern of results as for behavioural performance, depending on the measure. For example, longer viewing of the moved object and the previous position of the moved object were hypothesised to correlate with memory performance and show a similar condition split. Inversely, shorter viewing on stationary objects at retrieval was expected to track group differences via memory performance.

Similarly, evidence suggests that when gathering visuospatial information, participants can fixate on the 'centre of gravity' of the objects, or roughly the average of their spatial positions, hereafter referred to as the 'centroid' (Vish-

wanath and Kowler, 2003; McGowan et al., 1998; Zelinsky et al., 1997). Fixations on the centroid have been shown to correlate with task performance in visual search tasks (Venini et al., 2014), and have been linked to optimal visual information gathering (Najemnik and Geisler, 2005; Zelinsky, 2012). There is evidence for a reduced effect of this in AD (Rösler et al., 2005), although an examination of centroid fixations and memory performance has not been undertaken, to the best of my knowledge. However, improved multiple object tracking across viewpoint shifts has been linked to centroid fixations (Huff et al., 2010). Therefore, a corollary eye movement prediction was that MCI+ participants would have fewer fixations on the centroid compared to control participants.

Viewing of the centroid and the previous position of the moved objects were not measured directly. This is because the world-space position of these locations changed for Rotate trials. Consequently, participants may have fixated on the previous position based on the pre- or post-rotated configuration. The comparability of these to other conditions was difficult to compare across conditions in these cases. Furthermore, the centroid of the object configuration changes from encoding to retrieval with the shift in the moved object. Therefore, I decided to include viewing of the table itself as an indirect measure of these areas of interest (AOIs). As the direction of hypotheses for previous position and centroid viewing was the same, then the direction for table viewing was also the same: greater table viewing was predicted to be associated with higher memory performance and younger, healthier participants.

Viewing behaviour was quantified using fixation-based measures per viewing period. For example, the proportion of time spent fixating on an AOI ('dwell proportion'), or the proportion of the number of fixations on that AOI ('proportion of fixations'). A further set of these measures involves calculating re-fixations or re-visits on AOIs. This requires isolating the external fixations (fixations between AOIs, rather than within) and calculating dwell proportion ('re-dwell') or

counting fixations ('re-visits') from all but the first fixation on that AOI. This chapter will predominantly describe measures derived from fixation *time* i.e. dwell proportions, with fixation count measures provided in the Appendix (Chapter B).

A key contribution of this work was to examine fixation-derived measures at the retrieval phase of the trial (second viewing). However, we can also test memory-related eye movement effects by examining the *change* in viewing behaviour from first and second viewings, such as the increase in dwell proportion ('dwell increase') on the moved object. Fixation measures derived from this difference were primarily examined for the moved object and stationary objects.

Eye movement hypotheses at encoding. In the first viewing, healthy participants with higher memory performance were expected to utilise memorisation strategies (Hilton et al., 2020). A hypothesised effect of this on eye movements was that Younger, Older and MCI- participants would have a more predictable fixation pattern than MCI+ participants, as measurable by decreased transition entropy (Lucas et al., 2018; see Section 2.6.1).

Additionally, MCI+ participants were hypothesised to exhibit reduced scanpath reinstatement due to damage in the medial temporal lobe (see Section 1.4). To measure this, the MultiMatch 'shape' measure was used to compare temporal sequences of world-space saccade vectors from encoding to retrieval (Jarodzka et al., 2010; Dewhurst et al., 2012). However, the measure was not used to compare table-rotation conditions (Stay-Rotate, Walk-Rotate, Teleport-Rotate) due to the difference in world-space object positions from encoding to retrieval (further explanation in Section 2.6.1 in Materials and Methods).

Finally, healthy participants were expected to spend more time looking at the centroid or space between objects at encoding, for the same reasons as at retrieval. Again, this was measurable by increased fixation time on the table.

Confidence hypotheses. Confidence ratings were hypothesised to correlate with task performance in healthy groups but not in MCI+ participants due to memory-monitoring impairment. This was measured by (a) how well task performance predicted confidence, and (b) by the difference in confidence between correct and incorrect trials. Confidence-condition interactions were expected to follow memory performance for healthy adults.

4.3.2 Statistical and analytical approach

Mixed-effects two-way ANOVAs were run for each measure to test for an effect of group (between-subjects factor) or condition (within-subjects factor) or an interaction between the two, with random effects of participants on intercepts. Post hoc testing of interaction effects were examined by running the same tests on two subsets of groups: healthy ageing (Younger and Older) and MCI biomarkers (MCI- and MCI+). Results of these effects are provided in the main body of text.

More granular post hoc comparisons were made within groups and conditions alongside visualisations with bar-charts. The effect of group on each dependent variable was tested within each condition using one-way ANOVAs, with pair-wise comparisons made using independent-samples *t*-tests if the dependent variable was at the participant level (e.g. percentage of correct trials) or hierarchical mixed-effects linear models for trial-level data (e.g. confidence rating, or dwell proportion on the moved object at retrieval), accounting for random effects of participants. Similarly, the effect of condition on each dependent variable was tested within each group using repeated measures ANOVAs and dependent-samples *t*-tests for pairwise comparisons at the participant level, or hierarchical mixed-effects models for trial-level data.

Assumptions of normality for these tests were assessed by visual inspection and Shapiro-Wilk tests (Shapiro and Wilk, 1965). Log-transformed data or non-parametric tests were used if normality was violated. For example, some eye-tracking measures did not satisfy assumptions of normality but rather

followed a log-normal distribution. Accordingly, dwell proportions were transformed using the natural logarithm of the measures in these cases, as judged by visual inspection of distributions. To avoid infinite values from log-transforming zero values, the minimum value was added to all values before the transformation. The post-transformation minimum value was also added to all measure values to keep 0 at 0 and maintain a positive direction (therefore a higher log of dwell proportion means higher raw dwell proportion). Where log transformations have been made in subsequent sections, the y axis raw values are shown on the right-hand side of the plot for translation of log values.

To test for significant associations between variables, measures were aggregated to the participant level of granularity before running univariate linear regressions. The relationship between task performance and more granular measures were also investigated by testing the effect of trial correctness on these measures using the same statistical tests as above.

Statistics were run using the pingouin (Vallat, 2018), statsmodels (Seabold and Perktold, 2010) and scipy (Virtanen et al., 2020) Python packages. In-text ANOVA statistics are reported from pingouin model outputs with *F*-values, *p*-values, generalised eta-squared η_g^2 for within-subject effect sizes, and partial eta-squared η_p^2 for between-subject effect sizes. Details of post hoc pairwise statistical tests were not included in-text to improve readability, but are marked on relevant figures.

4.3.2.1 Factorial comparisons of spatial consistencies

An alternative way to compare conditions in this spatial updating paradigm—introduced by Burgess et al., 2004—involves treating the six conditions as variations along three spatial factors. These are described and explained in this section, where the reader is encouraged to refer to Figure 4.5 for different factorial designs and predicted performances.

As mentioned previously, different conditions vary by whether the object po-

sitions at retrieval are consistent with an egocentric *visual* snapshot of the configuration at encoding: if the first-person perspective of the object configuration is the same between first and second viewings, an egocentric representation of the spatial arrangement can be used to solve the task.

Alternatively, the configuration at retrieval may be consistent with an updated representation via *self-motion*. If the participant stays at the first viewpoint and the table does not rotate (Stay-Still), or the participant walks to a new viewpoint and the table is still (Walk-Still), then the configuration of the objects at retrieval is consistent with how the participant has moved in the environment.

In replication of Simons and Wang's early experiment, the Stay-Still condition was consistent with both *visual* and *self-motion* and therefore should have provided the greatest memory benefit in healthy participants. The Stay-Rotate condition was inconsistent with both, and therefore should have elicited the lowest performance, as in previous experiments (Simons and Wang, 1998; Burgess et al., 2004). Walk-Still was consistent with *self-motion* but not *visual*, and Walk-Rotate is consistent with *visual* but not *self-motion*.

Previous results have shown a benefit to memory performance from consistency of either spatial factor in healthy younger participants with a 2x2 factorial design (Simons and Wang, 1998; as in Figure 1.5 in Chapter 1). Accordingly, results were statistically analysed using this approach in addition to the tests described in the previous section. By this method, we would expect MCI+ participants to show no effect of *self-motion* consistency on task performance. As mentioned previously, this would result in a similarly low performance in Walk-Still as the Stay-Rotate condition, unlike healthy participants. In accordance with this model of the paradigm, MCI+ participants should also show reduced performance in the Stay-Still condition, which is consistent with self-motion updating. This is contradictory to the previously mentioned hypothesis that the Stay-Still performance in MCI+ participants will not be worse than MCI- or Older groups due to the availability of the egocentric *visual* snapshot. Note that a

deviation from the original paradigm was introduced here: participants were not required to walk halfway and back to the original viewpoint in Stay-Still trials. The consequences of this are discussed at the end of the chapter (also in the general discussion in Chapter 7).

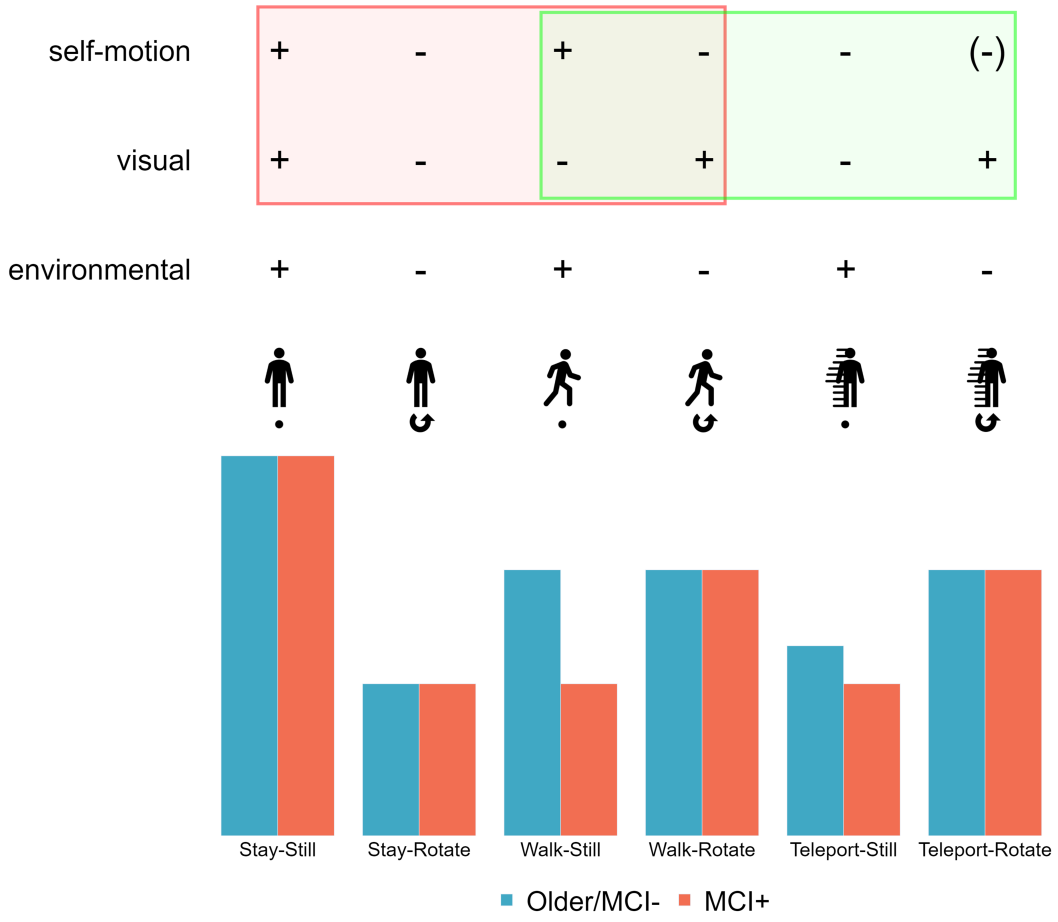


Figure 4.5. Hypotheses by condition, split by MCI+ and direct control groups. The top plot shows consistency of each condition by different spatial factors, with coloured boxes highlighting different factorial designs. Red is an approximate replication of Simons and Wang's original study (Simons and Wang, 1998), and green is the alternative test of self-motion benefit using the teleport conditions. The Teleport-Rotate condition is likely inconsistent with self-motion, as described in the main text, which can be tested by the green box 2x2 factor design. The bottom bar chart shows hypothesised relative scores for memory-related measures (i.e. task performance and memory-related eye movement measures), comparing MCI+ participants to healthy controls.

Teleportation and self-motion consistency. The teleporting conditions were included to disrupt self-motion. However, a question remains whether this form of movement can be used to assume *self-motion* consistency under certain conditions. As mentioned in Section 4.3.1, if we assume teleporting does *not* involve

self-motion updating of encoded representations (discussed further down), then the Teleport-Still condition is inconsistent with *visual* and *self-motion* factors and should elicit reduced performance in healthy participants compared to Walk-Still, which is consistent with *self-motion*. Additionally, Teleport-Rotate trials are consistent with *visual* representations similar to Walk-Rotate. However, if teleporting is not self-motion, then this condition also provides consistency with *self-motion* (unlike Walk-Rotate) because the participant has not walked to a new viewpoint yet is viewing the objects from the same egocentric snapshot as at encoding. Under this argument, comparing Walk and Teleport conditions provides a second 2x2 factor design for testing the effects of *self-motion* and *visual* consistency.

However, by this model we would hypothesise a similar performance in Teleport-Rotate as in the Stay-Still condition. This seems unlikely for two main reasons.

Firstly, the act of teleporting to a new location may be disorienting, particularly because participants were required to physically rotate towards the table after teleporting in a straight line (Section 2.2). In fact, this may have had an effect on performance by itself. Indeed, we can test whether there was a general performance-reducing effect of teleporting overall by comparing Walk and Teleport conditions: if teleporting affects performance per se, then we would expect all Teleport conditions to be lower than Walk trials, regardless of table rotation. However, the alternative hypothesis, that teleporting only reduces performance with a viewpoint shift, would support this type of movement as a more pure disruption of self-motion processing, rather than a general disorienting factor.

Another reason why we would *not* expect Teleport-Rotate trials to elicit similar performance as Stay-Still is the consistency of the object configuration with the external environment. Indeed, a key difference between Teleport-Rotate and Stay-Still is that in the former, both the subject and table have moved in relation to the environment. An additional benefit of environmental (‘allocen-

tric') consistency compared to *visual* and *self-motion* has been demonstrated in healthy participants using a 2x2x2 factor design (Burgess et al., 2004; Figure 1.6 in Chapter 1). Therefore, we would not expect Teleport-Rotate performance to match Stay-Still and would hypothesise a non-significant effect of self-motion consistency when assuming that Teleport-Rotate provides consistency of the objects with self-motion.

It was not possible to replicate Burgess and colleagues's three-factor design to test for the effect of allocentric consistency using the conditions tested here, and indeed this was was not planned into the design of the task. A 2x2 factor analysis of *visual* and *environmental* consistency could be performed using Stay and Teleport conditions, but the Stay condition would require some control for the potentially confounding act of teleporting. Future work could investigate this in early AD participants by rotating the environment as well as the subject and the table.

In sum, the effect of self-motion and visual consistency with the object configuration on task performance was tested with two separate 2x2 factorial comparisons: one using Stay and Walk conditions in replication of previous studies, and one using Walk and Teleport trials. The latter combination was actually expected to *not* match the first due to the role of allocentric consistency and the disorienting effect of teleporting compared to staying at the same viewpoint. However, a main effect of teleporting was also tested for, with the expectation that teleporting reduced performance more in Teleport-Still than Teleport-Rotate due to availability of egocentric *visual* consistency.

These comparisons were performed for two subsets of groups each: healthy ageing (Younger and Older) and MCI groups (MCI- and MCI+). Effects of self-motion consistency, visual consistency, their interaction with each other, and their interactions with group were tested using three-way mixed-effects ANOVAs (three-way interaction effects are reported from statsmodels output using model

β-coefficients and *p*-values). The main dependent variable of interest was task performance (percentage of correct trials). However, the same analyses were run for confidence rating and dwell proportion on the moved object, as these were expected to closely relate to task performance.

4.4 Results

The following sections include many statistical and graphical results. To aid understanding, key findings are summarized at the start of each section, though an overview of all findings is provided in the final part of the chapter.

4.4.1 Memory performance

Summary of key findings

- Unexpectedly, MCI+ and MCI- participants did not show significant differences in any condition on percentage of correct trials.
- Older adults performed significantly better than MCI+ participants in several conditions, with the greatest difference in Stay-Still trials.
- There was evidence of a reduced effect of self-motion consistency for MCI+ participants, consistent with hypotheses.
- Younger adults scored higher than older adults in all conditions.
- There was evidence of a general performance-reducing effect of teleporting in healthy groups, contrary to hypotheses.
- Task performance in the Stay-Still condition was significantly greater than all other conditions for Younger, Older and MCI- groups.
- MCI- participants performed equally low in all conditions other than Stay-Still, although all groups scored above chance in all conditions.

Figure 4.6 presents results of task performance between groups and conditions as measured by percentage of trials where the moved object was correctly selected. There was a significant main effect of group ($F(3,\ 84) = 14.9,\ p <$

$.001)$ on percentage of correct trials. Figure 4.6a visualises pairwise differences between groups when all conditions were combined, showing significantly greater performance in the Younger group than all others across the whole task, with a weaker difference between Older and MCI+ groups. No significant difference between MCI- and MCI+ groups was observed.

No group-condition interaction effects were found when including all groups $(F(15,\ 420) = 1.4,\ p = .14)$ or just Younger and Older groups $(F(5,\ 320) = 0.7,\ p = .59)$. This latter result supports the hypothesis of a generally lower performance with age in spatial memory, with no specific impairment in allocentric memory.

However, a significant interaction of group and condition was found when including only MCI- and MCI+ participants $(F(5,\ 100) = 2.6,\ p = .03)$. Looking between groups within conditions (Figure 4.6b), this may be explained by non-significantly greater MCI- mean scores in all conditions except for the Walk-Rotate condition where the effect was reversed. Otherwise, pairwise significant differences predominantly consisted of greater performance in the Younger group than other groups. In addition, a significant difference between Older and MCI+ groups was observed in the Stay-Still, Walk-Still, Stay-Rotate and Teleport-Rotate conditions, and a significant difference between Older and MCI- groups in the Walk-Rotate condition only. No significant pairwise differences were observed between MCI- and MCI+ groups. This pattern of results is contrary to hypotheses in that the only significant differences between MCI+ and either MCI- or Older participants was expected in the Walk-Still and Teleport-Still conditions, with all other conditions showing similar levels of performance between these groups.

All participants performed statistically significantly better than chance in all conditions, tested with one-sample *t*-tests against a score of 0.25 $(p < .001)$.

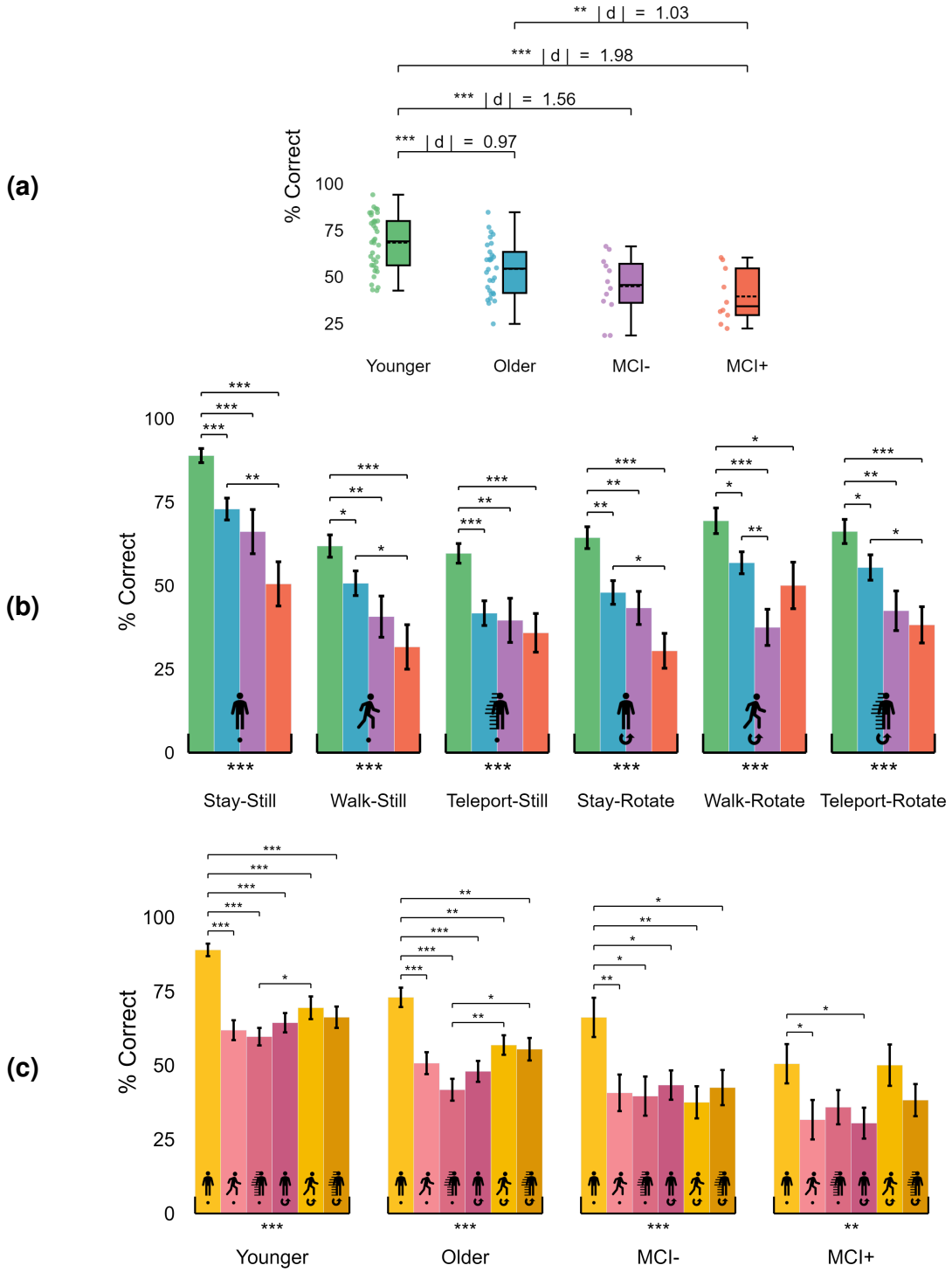


Figure 4.6. Percentage of correct trials. (a) Total percentage of correct trials across all trials between groups. Significance brackets include absolute Cohen's d effect sizes. (b) Within-condition group comparisons of task performance. Main effect significance is represented by brackets at the base of bar clusters. Brackets above bars show pairwise significance. Error bars are +/- standard error of the mean. (c) Within-group condition comparisons of task performance, with between-condition pairwise comparisons. Yellow-toned bars are conditions with egocentric (visual) consistency, pink-toned bars are conditions with a shifted view of the objects.

Between-condition effects and factorial comparisons of spatial consistencies. There was a significant main effect of condition on percentage of correct trials $(F(5,\ 420) = 43.4, p < .001)$. Examining effects between conditions within groups (Figure 4.6c), significant post hoc main effects of condition were found within all groups. Pairwise significant differences between Stay-Still conditions and all other conditions were observed within Younger, Older and MCI- groups, showing increased performance in this condition. In addition, the Younger group showed a decrease in performance for the Teleport-Still condition compared to the Walk-Rotate condition. The Older group additionally showed decreased performance in Teleport-Still compared to Teleport-Rotate trials. The MCI+ group scored unexpectedly high on average in the Walk-Rotate condition compared to other conditions, although this was not statistically significant by pairwise analysis.

When isolating Younger and Older groups, significant main effects of visual consistency $(F(1,\ 29) = 32.3,\ p < .001)$, self-motion consistency $(F(1,\ 29) = 16.6,\ p < .001)$ and an interaction between the two were found $(F(1,\ 29) = 13.3,\ p = .011)$ based on Stay and Walk conditions. This interaction effect is likely explained by the higher scores in the Stay-Still condition than other conditions and the relatively similar performance in the Stay-Rotate as the Walk-Still condition in both groups. This is contrary to hypotheses and previous results (Simons and Wang, 1998; Burgess et al., 2004), which predicted greater performance in Walk-Still compared to Stay-Rotate due to self-motion consistency. Indeed, the main effect of self-motion consistency seems to be driven by the Stay-Still condition alone, hence the significant interaction between self-motion and visual consistency.

No significant interactions between group and either visual consistency $(F(1,\ 29) = -3.84,\ p = .45)$, self-motion consistency $(F(1,\ 29) = -5.27,\ p = .30)$ or their interaction $(F(2,\ 29) = 8.70,\ p = .22)$ were found for healthy groups. This suggests no difference in how these spatial consistencies af-

fect different healthy age groups, which is consistent with the hypothesised condition-invariant performance decrease in Older participants compared to Younger.

When examining visual and self-motion consistency using the alternative approach of Walk and Teleport conditions, no effect of self-motion consistency was found ($F(1,\ 29) = 6.77,\ p = .16$), contrary to when using Stay and Walk conditions. As previously described, this may suggest that this combination of conditions is not appropriate for studying self-motion consistency, rather than indicating a contradictory absence of effect. In support of this, a significant interaction between visual and self-motion consistency was found in healthy groups when isolating these conditions ($F(1,\ 29) = 10.4,\ p = .039$), which is likely explained by the relatively lower performance of both Teleport-Still and Teleport-Rotate to their Walk counterparts. If Teleport-Rotate was consistent with self-motion, we would expect a performance boost to Teleport-Rotate over Walk-Rotate. Instead, a significant main effect of teleporting was found in both healthy groups ($F(1,\ 29) = 8.96,\ p = .011$) with no interaction with table rotation ($F(1,\ 29) = 7.55,\ p = .13$) or between groups ($F(1, 29) = 6.77,\ p = .16;\ F(1,\ 29) = 8.56,\ p = .21$). These results suggest that, for both age groups, there was a general performance-reducing effect of teleportation, regardless of the availability of egocentric strategies. This is contrary to the hypothesis that teleportation will not provide disorientation beyond disruption to self-motion.

A significant main effect of visual ($F(1,\ 9) = 19.57,\ p = .002$) but not self-motion consistency ($F(1,\ 9) = -4.22,\ p = .51$) or their interaction ($F(1,\ 9) = -0.68,\ p = .94$) was found on task performance for MCI groups when isolating Stay and Walk conditions. In addition, significant interactions between group (biomarker status) and (a) the effect of visual consistency

$(F(1,\ 9) = -25.38,\ p = .004)$, and (b) the interaction effect between visual and self-motion consistency $(F(1,\ 9) = 32.0,\ p = .010)$ were found. These results suggest that the effect of visual consistency on performance varied by biomarker status, as did the interaction between visual and self-motion consistency. In support of this, Figure 4.6c shows that MCI- participants performed much higher in Stay-Still trials (visual and self-motion consistency) than all other conditions, which were all relatively equal. Therefore, only when both visual and self-motion consistency were present was there a performance boost for biomarker-negative participants. In contrast, MCI+ participants had similarly high performance in both Stay-Still and Walk-Rotate conditions (visual consistency) but similarly low performance for Walk-Still and Stay-Rotate (visual inconsistency). This suggests that there was no performance-enhancing effect of self-motion consistency for MCI+ participants, and that there was a general performance-enhancing effect of visual consistency that even surpassed the MCI- group. This is somewhat consistent with hypotheses, except that the MCI- group were expected to show an effect of self-motion consistency, and were not expected to be outperformed by MCI+ in any condition.

No main effect of teleporting was found in MCI participants $(F(1,\ 9) = 4.22,\ p = .51)$, with no interaction with biomarker status (group) $(F(1,\ 9) = -5.33,\ p = .54)$. This suggests there was no general performance-reducing effect of teleporting compared to walking for MCI- or MCI+ participants.

4.4.2 Confidence

Summary of key findings.

- MCI+ participants had less difference in confidence ratings between correct and incorrect trials than Older and Younger groups, but not MCI- participants.
- Younger and Older groups' overall task performance significantly predicted their mean confidence ratings, but the same did not hold for MCI groups.
- A significant confident-reducing effect of teleporting was found in healthy groups, consistent with task performance.
- Younger adults were significantly more confident than Older and MCI+ participants, but not MCI- participants, across all conditions.

Figure 4.7 shows group differences in confidence rating per trial. Significant main effects of group $(F(3,\ 86) = 4.8,\ p = .004)$ and condition $(F(5,\ 430) = 51.4,\ p < .001)$ were found on confidence scores. Figure 4.7a shows greater overall mean confidence ratings in the Younger group than Older and MCI+ groups, but not MCI- participants.

A significant group-condition interaction was found for healthy ageing groups $(F(5,\ 330) = 3.9, p = .002)$ but not MCI biomarker groups $(F(5,\ 100) = 1.1,\ p = .38)$, suggesting that differences in confidence between conditions varied by healthy age group, but not MCI groups. Within conditions (Figure 4.7b), significant main effects of group were found in all conditions, mainly driven by significantly higher confidence in Younger participants compared to Older and MCI+ participants. MCI- participants were not significantly less confident than

Younger participants in any condition.

Across conditions within groups (Figure 4.7c), main effects of condition were found in all groups. Younger, Older and MCI- groups had significantly higher confidence in the Stay-Still condition than all other conditions. Younger participants also had significantly increased confidence in the Walk-Rotate condition compared to Walk-Still and Teleport-Still conditions. This pattern of pair-wise differences in the Younger group is comparable to that for percentage of correct trials. This consistency between task performance and confidence rating across conditions was also broadly found for other groups: in the Older group, significantly lower confidence was found in the Teleport-Still condition compared to the Walk-Rotate trials, similar to task performance. Likewise, no significant differences in confidence were observed in the MCI- group between conditions other than increased Stay-Still confidence. MCI+ participants had higher confidence ratings in Stay-Still and Walk-Rotate conditions compared to Walk-Still, Teleport-Still and Teleport-Rotate conditions, reflecting higher task performance in these conditions.

Factorial comparisons of spatial consistencies. A significant main effect of visual consistency $(F(1, 29) = 18.5, p < .001)$, self-motion consistency $(F(1, 29) = 4.49, p = .041)$ and their interaction $(F(1, 29) = 14.3, p < .001)$ was found for healthy age groups when isolating Stay and Walk conditions only. Additionally, a significant three-way interaction between group, visual consistency and self-motion consistency was found $(\beta(1, 29) = 0.928,\ p = .004,\ 95\%CI[0.29, 1.56])$. These results suggest that the effect of visual consistency on confidence varied by self-motion consistency, and that this interaction was different for each age group. Figure 4.7c indicates that Younger participants had a greater increase in confidence from the Stay-Still condition versus other conditions compared to the Older group.

A significant main effect of visual consistency $(F(1, 9) = 7.02,\ p = .026)$

was found for MCI biomarker groups when isolating Stay and Walk conditions. A main effect of self-motion consistency may be present but did not reach statistical significance $(F(1,9) = 5.62,\ p = .069)$. Likewise, an interaction between visual and self-motion consistency did not reach statistical significance $(F(1,9) = 7.4,\ p = .09)$. These results suggest that visual consistency provided a boost to confidence in these groups, but self-motion provided a weak effect at best, which varied weakly by visual consistency. These effects did not vary significantly by group $(p > .05)$ and there was no statistically significant overall main effect of group $(\beta(1,29) = 0.93,\ p = .32,\ 95\%\ CI\ [-0.90,\ 2.76])$ despite the consistently lower mean confidence for MCI+ compared to the MCI- group. Figure 4.7c suggests that there was a boost to confidence in the Stay-Still condition for MCI- participants only, but this effect was not statistically greater than MCI+ participants.

A significant main effect of teleporting on confidence was found for healthy participants only $(F(1,29) = 8.94, p = .005)$, suggesting that the act of teleporting reduced confidence in task performance compared to walking. This effect did not vary by group $(\beta(1,29) = 0.074,\ p = .75,\ 95\%\ CI\ [-0.39,\ 0.53])$.

Confidence associations with task performance. Figure 4.8 visualises the relationship between performance and confidence within groups. Figure 4.8a shows results of linear regressions between performance and overall mean confidence rating, split by group. Mean confidence rating significantly predicted percentage of correct trials in both Younger and Older conditions but not MCI- or MCI+ groups. This means that no evidence could be found for a relationship between task performance and confidence for MCI groups based on this sample size.

Figure 4.8b shows confidence ratings when split by correct and incorrect trials. There was a strong main effect of correctness on confidence $(F(1,95) = 161.2, p < .001, \eta_g^2 = .75)$, with post-hoc tests showing that all groups had signif-

icantly lower confidence ratings in incorrect trials than correct trials $(p < .001)$. There was also a small main effect of group ($F(3, 95) = 4.07,\ p = .009,\ \eta_p^2 = .11$), with post hoc pairwise tests showing significantly lower confidence in Older than Younger participants $(t(51.8) = -3.19,\ p = .002$; degrees of freedom adjusted using Greenhouse-Geisser correction$)$, but no other pairwise significance $(p > 0.05)$.

In incorrect trials only, MCI groups had higher average confidence ratings than healthy groups, although this was not statistically significant $(p > .05)$. There was a moderate significant interaction between group and correctness $(F(3, 95) = 17.7, p < .001, \eta_p^2 = .25)$, reflecting the greater reduction in confidence in incorrect trials for healthy groups compared to MCI groups (4.8b).

To quantify this interaction at the participant level, Cohen's *d* effect sizes were calculated for the difference between correct and incorrect confidence per participant. Figure 4.8c visualises the difference in confidence rating by correctness, representing the increase in confidence for correct trials compared to incorrect trials. Younger participants had significantly greater confidence differences than all other groups here, with an additional significant pairwise difference between MCI+ and Older participants, but not MCI- participants. This suggests that MCI+ participants had more decoupling between confidence and task performance than Older and Younger groups, but not MCI- participants.

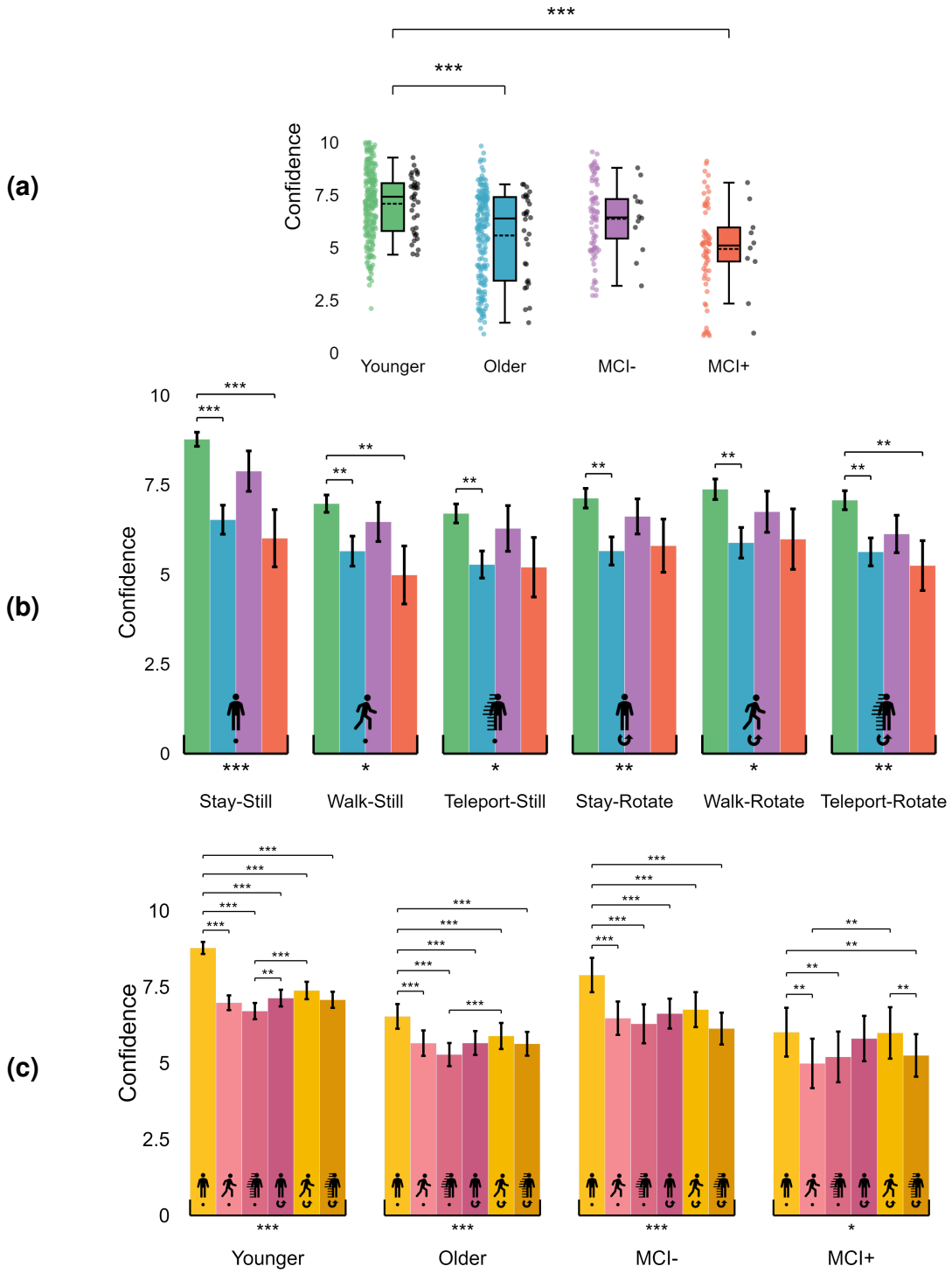


Figure 4.7. Confidence ratings. (a) Overall confidence between groups. Coloured points represent mean confidence by condition (6 points per participant); black points represent total mean confidence by participant (1 point per participant). Significance brackets are from pair-wise hierarchical linear models. (b) Within-condition group comparisons. ANOVA main effect significance is represented by brackets at the base of each cluster. Pairwise differences are represented by brackets above bars. Error bars are standard error of the mean. (c) Within-group condition comparisons.

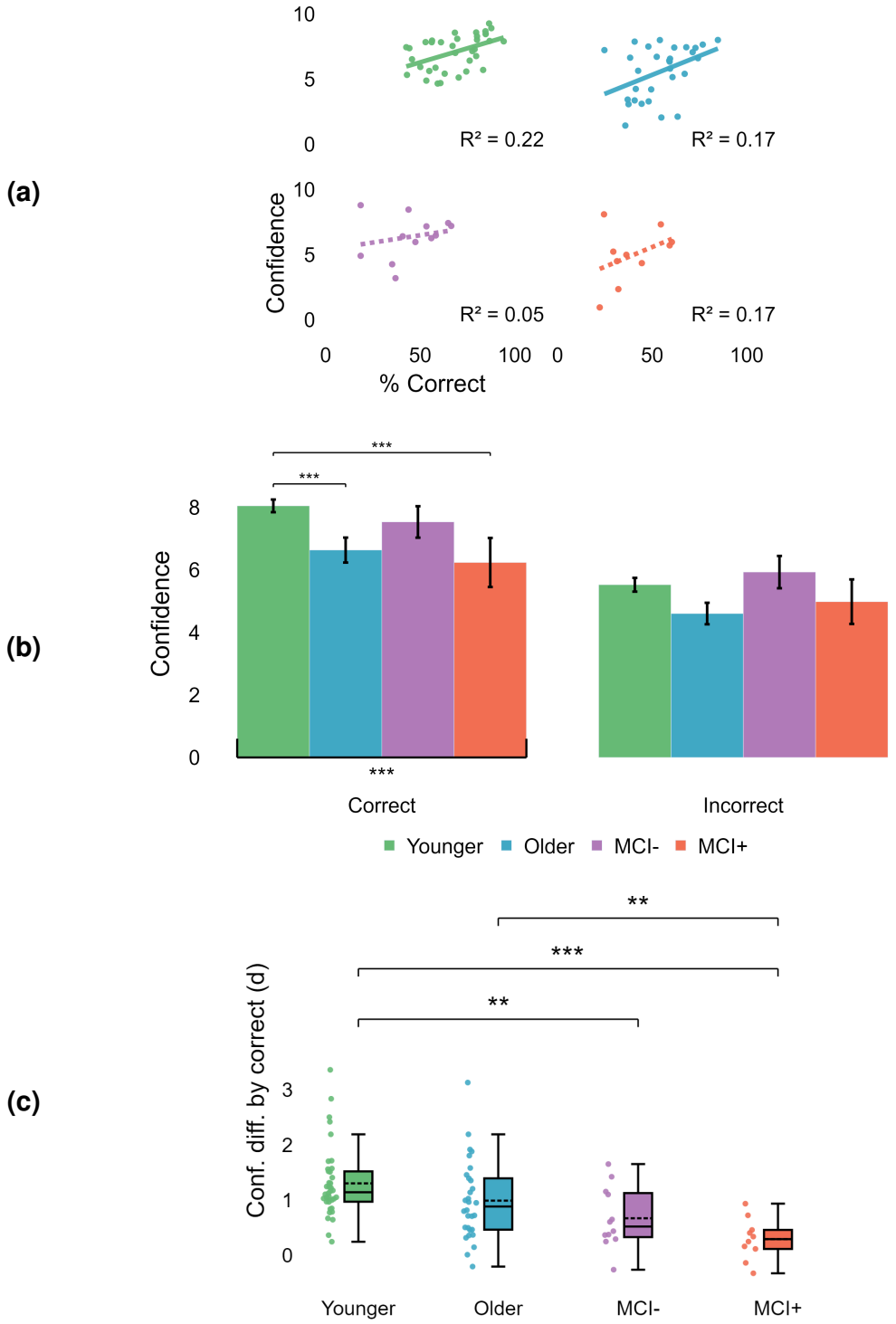


Figure 4.8. (a) Scatter plots of confidence by percentage of correct trials, compared per group. Lines are linear models; solid lines are statistically significant, $p < .05$; dashed lines are not. (b) confidence rating split by group, within correct and incorrect trials. (c) Cohen's d per participant of the confidence rating difference in correct versus incorrect trials, split by group

4.4.3 Eye movements

Although eye-tracking data were collected across the entirety of each participant's time wearing the head-mounted display, the current analysis only examines eye movement patterns during encoding and retrieval phases. These were designed to be compared and contrasted, lasting the same length of time (7 seconds). Accordingly, these phases were the subject of the hypotheses outlined earlier (Section 4.3.1).

As mentioned in Section 2.6.1 on eye movement measures, there are a number of approaches to analysing eye movements. This section focuses on testing hypotheses from this chapter (Section 4.3.1), comparing measures at the viewing- or trial-level, such as dwell proportion on the moved object at retrieval. Key eye movement measures were also compared between correct and incorrect trials per group to examine their associations with task performance. Analysis was extended to include post hoc comparisons based on relative viewing patterns between areas of interest (AOIs).

Further exploration of eye behaviour is detailed in Chapter 5 (also see Appendix A).

4.4.3.1 Eye movement measures at retrieval

Summary of key findings.

- MCI+ participants had significantly less dwell time on the stationary objects than MCI- *and* Older participants, consistent with hypotheses. However, this effect was unexpectedly strongest in the Stay-Still condition.
- Similarly, MCI+ participants had significantly less dwell time on the table overall, but particularly in the Stay-Still condition when compared to both MCI- and Older groups.
- Younger participants fixated on the table significantly more than all other groups.
- No statistically significant differences between groups were observed in fixation time on the moved object, contrary to hypotheses.
- Younger, Older and MCI- groups had significantly greater fixation time on the moved object in the Stay-Still condition than all other conditions, consistent with task performance and confidence ratings.
- Teleporting did *not* effect fixation time on the moved object in healthy groups, contrary to task performance and confidence ratings.

Fixations on the moved object. The first memory-related eye movement measures to examine were derived from fixations on the moved object. Figure 4.9 shows results of group differences in log-transformed dwell proportions (log-dwell) on the moved object. Contrary to hypotheses, no significant differences

between groups were found. No main effect of group was found when combining all conditions ($F(3, 94) = 1.3, p = .28$), although from pairwise comparisons the Older group had significantly greater log-dwell than MCI- and MCI+ groups (Figure 4.9a). Still, this suggests no overall detectable difference between groups in the proportion of time spent fixating on the moved object at retrieval.

A significant main effect of condition was found ($F(5, 470) = 11.9, p < .001$) but no significant interaction between group and condition when including all groups ($F(15, 425) = 1.2, p = .24$), healthy Younger and Older groups only ($F(5, 370) = 1.1, p = .36$), or MCI groups only ($F(5, 100) = 1.6, p = .16$). These results suggest an effect of condition on dwell proportion on the moved object that did not significantly vary between groups. When splitting by condition within groups (Figure 4.9b), a significant main effect of group was found for Walk-Still and Teleport-Rotate conditions. Only the latter showed a significant pairwise increase in log-dwell for the Older group compared to the Younger group. However, these post hoc effects may not be scientifically significant given the absence of a clear interaction between group and condition.

Differences between conditions within groups mainly showed increased log-dwell in the Stay-Still condition compared to other conditions in Younger, Older and MCI- groups (Figure 4.9c). These participants spent more time looking at the moved object in the Stay-Still condition compared to other conditions, consistent with task performance. The Older group additionally showed significantly higher log-dwell in Walk-Rotate and Teleport-Rotate conditions compared to Walk-Still and Teleport-Still, suggesting an effect of visual consistency (explored in the next section).

The MCI+ group was the only group to have no main effect of condition here; log transformed dwell proportions between conditions were not statistically different. This suggests that the MCI+ participants did not vary how much they looked at the moved object between different conditions. This is unexpected in relation to hypotheses, which predicted lower dwell on the moved

object in shifted viewpoint conditions (Walk-Still, Teleport-Still and Stay-Rotate) compared to conditions with availability of egocentric strategies (Stay-Still, Walk-Still, Teleport-Still).

The patterns of results were very similar for re-dwell (Supplementary Figure B.3), proportion of fixations (B.4), increase in proportion of fixations (B.5), dwell increase (B.6) and re-dwell increase (B.7) on the moved object.

Factorial comparisons of spatial consistency effects on moved object dwell proportion. A small significant main effect of visual consistency was found for healthy groups when isolating the Stay and Walk conditions $(F(1, 29) = 16.67,\ p < .001)$. However, no significant effect of self-motion consistency was found $(F(1, 29) = 2.11,\ p = .17)$. These results suggest that consistency of the object configurations with participant self-motion did not increase the log-transformed dwell proportion on the moved object at retrieval, contrary to hypotheses and results from memory performance and confidence ratings. However, a small significant interaction between visual and self-motion consistency $(F(1, 29) = 9.65,\ p = .004)$ alongside visual inspection of within-group condition differences (Figure 4.9c) suggests that there was greater log-dwell on the moved object in the Stay-Still condition compared to other conditions in the group. This interaction effect significantly varied between Younger and Older groups with a small effect size $(\beta(1, 29) = 0.07,\ p = .02,\ 95\%\, CI\ [0.009,\ 0.139],\ \eta_p^2 = .07)$, likely representing the greater difference between Stay-Still log-dwell and other conditions in the Younger group compared to the Older group. In contrast, the Older group had comparatively greater differences between Move-Rotate conditions and Move-Still conditions.

No significant main effect of visual consistency $(F(1, 9) = 4.20,\ p = .15)$, self-motion consistency $(F(1, 9) = 1.12,\ p = .74)$ or their interaction $(F(1, 9) = 0.14,\ p = .87)$ were found for MCI biomarker groups when iso-

lating Stay and Walk conditions. However, the interaction between visual and self-motion consistency significantly varied between MCI+ and MCI- groups $(\beta(1, 29) = 0.15,\ p = .007,\ 95\%\, CI\ [0.040,\ 0.264],\ \eta_p^2 = .13)$. This pattern of results suggests that visual and self-motion consistency did not have an effect on dwell proportion on the moved object, and only in the MCI- group did an interaction exist between the two. Indeed, MCI- participants had a significant difference between Stay-Still trials (visually and self-motion consistent) and all other conditions with no significant differences between remaining Stay and Walk conditions (Figure 4.9c). The MCI+ group on the other hand showed no significant main effect of condition on log-dwell on the moved object, suggesting no significant differences between conditions.

No significant main effect of teleporting was found in healthy groups $(F(1, 29) = 0.32,\ p = .56)$ or MCI groups $(F(1, 9) = 2.28,\ p = .16)$, with no interaction effects with group $(\beta(1, 29) = 0.014,\ p = .53,\ 95\%\, CI\ [-0.030,\ 0.058];$ $\beta(1, 29) = 0.015,\ p = .67,\ 95\%\, CI\ [-0.056,\ 0.675])$, suggesting that teleporting did not have an effect on the log-dwell proportion on the moved object in any group.

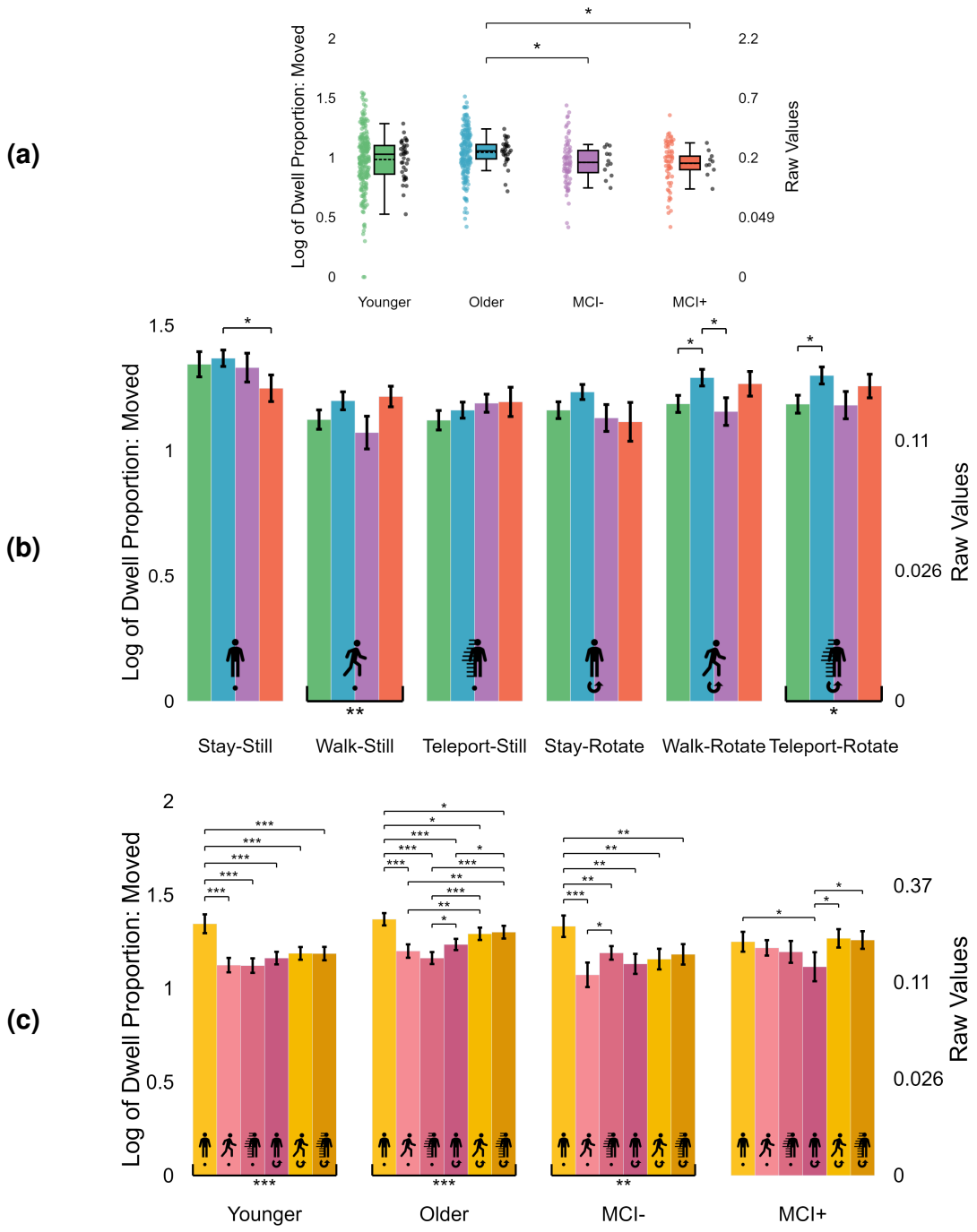


Figure 4.9. Dwell on the moved object. (a) Within-condition group comparisons of log-transformed dwell proportion. Note that the right-side y-axis shows the raw dwell proportions of the log-transformed y-axis on the left, which the data is plotted to. (b) Within-group condition comparisons of dwell proportion. (c) Within-condition group comparisons of dwell increase from first to second viewings. See Figure 4.7 for more details.

Fixation time on the stationary objects. Figure 4.10 shows log-transformed dwell proportions on the stationary configuration objects at the second viewing (retrieval). A significant main effect of group was found when combining all conditions ($F(3,\ 85) = 14.7,\ p < .001$), with pairwise results suggesting lower stationary dwell in the Younger group compared to Older and MCI+ participants (Supplementary Figure B.8a), and a greater log-dwell on the stationary objects in MCI+ compared to MCI- participants, supporting the hypothesis that MCI+ participants would show greater fixation time on the stationary objects than other groups.

Examining these differences by condition, the same pairwise effects were present in almost all conditions, with large differences between MCI+ and other groups in the Stay-Still condition (Figure 4.10a; $MCI+\ vs\ Older:\ t(38) = 4.11,\ p < .001,\ d = 1.50$; $MCI+\ vs\ MCI\text{-}:\ t(20) = 4.856,\ p < .001,\ d = 2.08$). This effect was even more pronounced between MCI+ and Older groups for the *re-dwell* on the stationary objects (the time spent revisiting these AOIs) in the Stay-Still condition (Supplementary Figure B.9b; $MCI+\ vs\ Older:\ t(38) = 5.16,\ p < .001,\ d = 1.88$; $MCI+\ vs\ MCI\text{-}:\ t(20) = 4.78,\ p < .001,\ d = 2.05$). Although these results support hypotheses that MCI+ participants would have greater fixation time on the stationary objects, it was the Stay-Still condition, rather than the Walk-Still condition, that showed the most difference in MCI+ participants based on these eye movements, contrary to hypotheses.

Further support for this is found in Figure 4.10c, which shows the decrease (first viewing minus second viewing) in raw dwell proportion on stationary objects between groups. The Stay-Still, Walk-Still, Stay-Rotate and Teleport-Rotate conditions had significant effects of group on this measure, with post hoc pairwise comparisons showing significantly less decrease of dwell proportion on the stationary objects in the MCI+ group compared to at least one other group. Again, the Stay-Still condition provided the strongest of these effects, with Younger, Older and MCI-groups showing significantly more decreased viewing

of stationary objects than MCI+ participants.

The above pattern of results was similar for proportion of fixations and re-fixations, as shown in the supplementary material (Appendix Section B.2.3). Between-condition effects are also described in the supplementary material (Figure B.8b), but essentially followed the inverse pattern of results described for dwell proportion on the moved object.

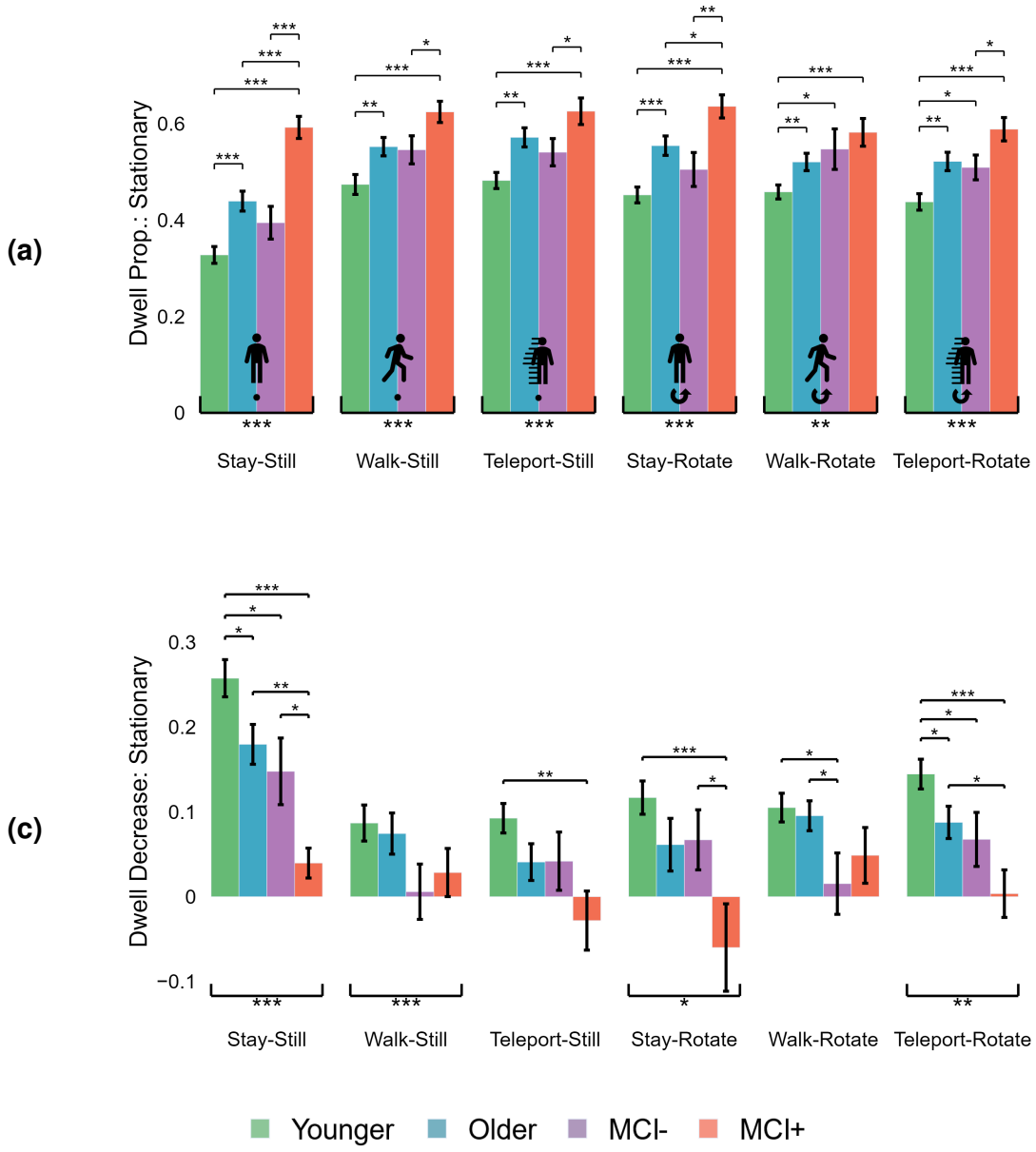


Figure 4.10. Dwell on the stationary objects. (a) Within-condition group comparisons of dwell proportion. (b) Within-condition group comparisons of *decrease* in dwell proportion from encoding to retrieval. See Figure 4.7 for more details.

Fixation time on the table. To capture fixations on the previous position of the moved object, the centroid of the object configurations, or on the spaces between objects, viewing behaviour on the table was examined.

Figure 4.11 shows a similar but inverted pattern between groups as Figure 4.10 on the stationary objects. A significant main effect of group was found for log-transformed dwell proportion on the table ($F(3,\ 85) = 11.2,\ p < .001$), with pairwise results suggesting greater table dwell in the Younger group compared to all other groups, and less table dwell in the MCI+ group compared to all other groups 4.11a. This is consistent with hypotheses that MCI+ participants would have reduced table viewing.

Looking within conditions, all conditions had main effects of group on table log-dwell. Pairwise differences showed less table dwell in the MCI+ group compared to all other groups in the Stay-Still condition, and less table dwell than MCI- participants in all conditions (Figure 4.11b). This is somewhat inconsistent with hypotheses that MCI+ participants would show reduced table viewing in Walk-Still and Teleport-Still conditions only.

Younger participants had significantly greater log-dwell on the table than Older participants in all conditions. Younger participants only had significantly greater log-dwell on the table compared to MCI- participants in the Walk-Rotate condition.

A significant main effect of condition was found on table log-dwell ($F(5,\ 425) = 4.2,\ p < .001$), although post hoc tests suggested no significant difference between conditions in the Older, MCI- or MCI+ groups (a direct between-condition comparison of dwell proportion on the table can be found in Supplementary Figure B.12). Younger participants had less table log-dwell in the Stay-Still condition than all other conditions, but the Older participants did not show this. Indeed, a significant interaction between healthy groups and condition was found ($F(5,\ 325) = 2.2,\ p = .05$), likely due to this effect. This is inconsistent with hypotheses in two ways: firstly, that table viewing did

not show a stronger condition-dependent effect, especially in the Older group; and secondly, that Younger participants had reduced table viewing in the condition that they show the greatest memory performance in. This latter effect will be explored further in Section 4.4.3.3 on eye movement associations with performance.

No significant interaction between MCI biomarker status and condition was found on table log-dwell ($F(5,\ 100) = 0.3,\ p = .88$), despite the significantly reduced log-dwell on the table in the MCI+ group for all conditions. This suggests that MCI+ participants have a reduced viewing of the table regardless of the consistency of the objects with egocentric, allocentric, or self-motion representations. Rather, a condition-invariant reduction in table viewing may exist in the MCI+ group compared to the MCI- group.

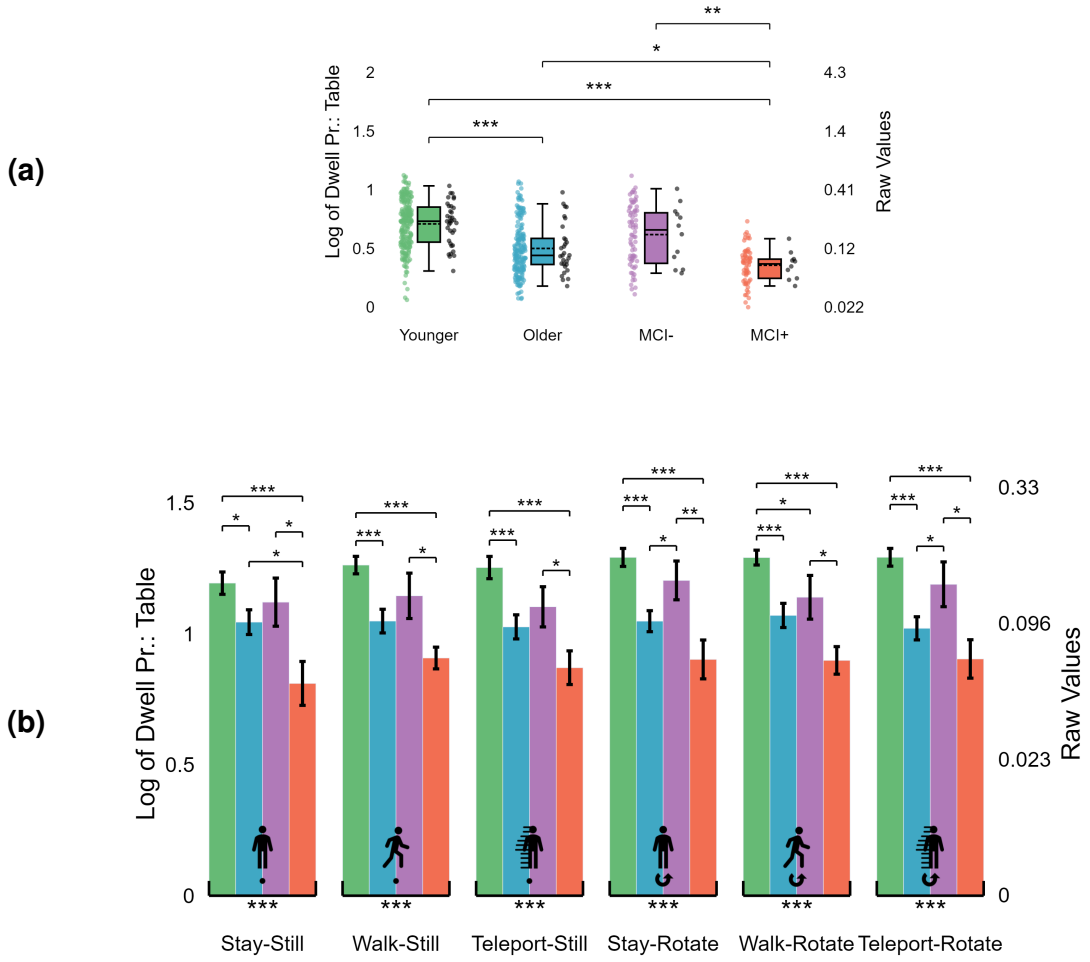


Figure 4.11. Log-transformed dwell proportion on the table. (a) Log-transformed dwell across all conditions between groups. (b) Within-condition group comparisons.

4.4.3.2 Eye movements at encoding

Summary of key findings.

- MCI+ participants had significantly lower MultiMatch shape similarity between encoding and retrieval scan-paths than all other groups in the Walk-Still condition. This suggests a reduced gaze reinstatement effect for these participants in this condition only. This is somewhat consistent with hypotheses, although the same effect was not observed in the Teleport-Still condition as expected.
- MCI+ participants significantly differed from MCI- participants in fixation time on the table at encoding, but not compared to Older participants.
- Younger adults fixated on the table significantly more than older adults during the encoding period.

Most dwell patterns at encoding were simpler to analyse because trials were condition-agnostic, therefore all trials could be used per group comparison. Main effects of condition were tested to confirm this.

Fixation time on the table. A significant main effect of group was found on log-transformed table dwell proportions at encoding $(F(3, 85) = 5.9, p = .001)$. Table viewing was expected to be higher for Younger and healthier groups, and indeed Figure 4.12a shows a significantly higher log-transformed dwell proportion on the table at encoding in Younger participants compared to Older and MCI+ participants, although not MCI- participants. Furthermore, a significantly greater table log-dwell was found for MCI+ participants compared to MCI- participants, but not Older participants.

No main effect of condition was found $(F(5, 425) = 0.2, p = .96)$.

Fixation time on the objects. A significant main effect of group was found on dwell proportions on all configuration objects at encoding $(F(3, 84) = 4.1, p = .01)$, with Younger participants spending less fixation time on the objects than Older and MCI+ participants (4.12b). MCI- participants also spent significantly less time fixating on the objects than Older participants. The pattern of group results for dwell proportion on the objects was approximately the inverse of dwell proportion on the table. The difference in statistical significance is likely explained by the log-transformation of the latter. Group differences in *re-dwell* on the configuration objects also followed a similar pattern (Supplementary Figure B.13).

These results suggest that MCI+ participants spent more time fixating and re-fixating on the objects during encoding than Younger and MCI- groups.

No effect of condition on either dwell $(F(5, 420) = 0.9, p = .49)$ or redwell $(F(5, 420) = 1.3, p = .27)$ on the objects was found.

Scan-path entropy. No significant main effect of group was found in scan-path entropy at encoding $(F(3, 85) = 0.8, p = .49)$, with no pairwise differences between groups (Figure B.14a). This suggests that groups did not differ in how random or constrained their viewing patterns were, contrary to hypotheses.

MultiMatch shape: scan-path vector sequence similarity. The final encoding-related metric was the MultiMatch 'shape' measure for estimating scan-path similarity between encoding and retrieval viewings based on the sequences of vectors in the two scan-paths. Like all MultiMatch measures, scores were bound between 0 and 1, with higher values indicating greater similarity between the two scan-paths being compared.

Figure 4.13 shows the results of this comparison between groups for each condition without table rotation (Stay-Still, Walk-Still, and Teleport-Still). A significant main effect of group was found on MultiMatch shape similarity between

encoding and retrieval scan-paths ($F(3,\ 85) = 7.1,\ p < .001$). Post hoc tests within conditions showed significantly less similarity between encoding and retrieval for the MCI+ group compared to the Younger group in all three conditions 4.13. In addition, the MCI+ group had significantly less similarity in the Walk-Still condition than all other groups. This suggests that MCI+ participants had less similarity in scan-paths between encoding and retrieval in the Walk-Still condition than other groups, consistent with hypotheses of reduced gaze reinstatement effects here.

No significant main effect of condition was found ($F(2,\ 170) = 0.9,\ p = .41$), suggesting that the condition (Stay, Walk or Teleport; table always Still) did not significantly affect vector sequence similarity between encoding and retrieval. However, a significant interaction between group and condition was found when including MCI biomarker groups only ($F(2,\ 40) = 5.2,\ p = .01$), with post hoc tests showing a greater difference in MultiMatch shape similarity scores between MCI- and MCI+ in the Walk-Still compared to the Teleport-Still and Stay-Still conditions. This is consistent with hypotheses that the Walk-Still condition would show specific MCI+ eye movement differences compared to the Stay-Still condition, although the Teleport-Still condition was expected to show similar results within groups.

No significant interaction between group and condition was found when including only healthy groups ($F(2,\ 130) = 0.5,\ p = .64$), suggesting that group differences in MultiMatch shape similarity between Younger and Older participants did not vary by condition.

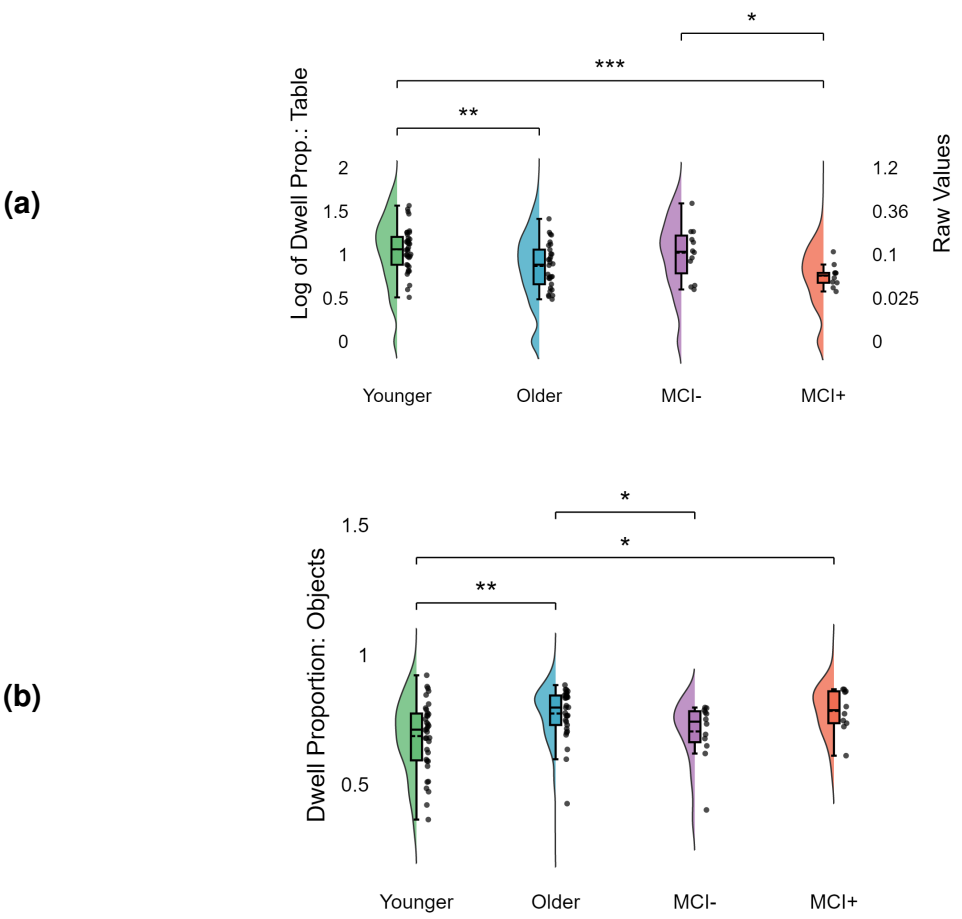


Figure 4.12. Raincloud plots (Allen et al., 2021) of group differences in viewing-level results at the encoding phase. Coloured density plots represent trial-level data, which were too numerous to show directly. Black data points represent participant-level averages, with box and whisker plots representing summary statistics of this data. Significance brackets are from hierarchical linear models. (a) Log-transformed dwell proportions on the table. (b) raw dwell proportions on configuration objects.

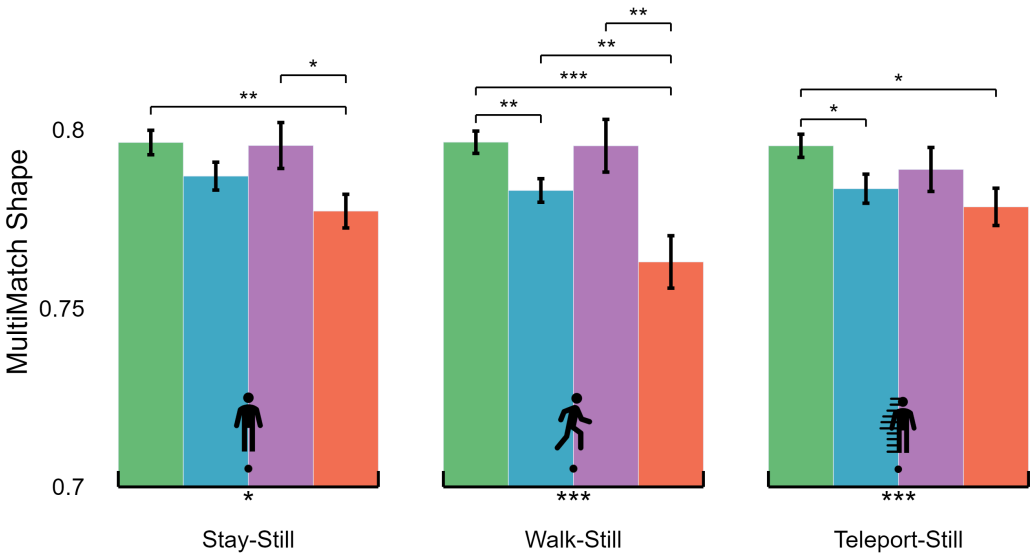


Figure 4.13. Grouped bar charts of MultiMatch shape similarity scores of encoding and retrieval vector sequences. A higher score indicates greater similarity between encoding and retrieval vector sequences. Rotation conditions are not shown due to the shift in world-space object positions with table rotation.

4.4.3.3 Eye movement associations with task performance

Summary of key findings.

- The decreased MultiMatch shape similarity between encoding and retrieval in Walk-Still trials for MCI+ participants may have been due to *correct* trials only.
- All groups looked at the moved object more, and the stationary objects less, in correct trials compared to incorrect trials, as expected.
- However, MCI+ participants did not show an interaction between this effect and condition, whereas other groups had even greater fixation time on the moved object, and even less on the stationary objects, in the Stay-Still condition compared to other conditions.
- Younger participants, unlike other groups, had greater fixation time on the table in *incorrect* trials, with an interaction between conditions not present in older adults.
- Older participants' fixation time on the table significantly predicted task performance.

Although associations are often examined with scatter plots and correlation analysis, this approach requires aggregation of data to a summary statistic per participant (e.g. the mean), losing most within-subject information. However, because trial correctness followed a binomial distribution, we can split data by correct/incorrect trials and compare eye movement associations against performance with more statistical power, especially given the nested structure of the data-set. Accordingly, the following sections show some key results of fixation

measures given a correctness split. Scatter plots with linear relationships between mean fixation metrics and overall task performance are also shown.

Dwell proportion on the moved object at retrieval. There was a strong main effect of trial correctness on dwell proportion on the moved object $(F(1, 85) = 239.3, p < .001, \eta_g^2 = .74)$. Figure 4.14a visualises the greater dwell proportion on the moved object in correct trials compared to incorrect trials (significance not marked on the figure).

For healthy Younger and Older groups, there was a significant interaction between correctness and group on moved object dwell proportion $(F(1, 65) = 239.3, p < .001, \eta_p^2 = .74)$, suggesting that the effect of age on moved object fixation time varied by trial correctness. Figure 4.14a shows dwell proportion on the moved object split by trial correctness, with significantly lower moved object dwell in Younger versus Older participants in correct trials only.

For MCI biomarker groups, there was no significant interaction between correctness and group on moved object dwell $(F(1, 20) = 1.38, p = .25, \eta_p^2 = .065)$, suggesting that the effect of biomarker status on moved object dwell did not vary by trial correctness. Figure 4.14a suggests that this is because there was little difference between MCI biomarker groups in their dwell proportion, regardless of trial correctness. However, there was a small significant interaction between trial correctness and condition in MCI groups $(F(5, 80) = 3.15, p = .012, \eta_g^2 = .041)$, suggesting that the effect of condition on moved object dwell proportion varied by trial correctness in these participants. A three-way ANOVA of condition, group and correctness on moved object dwell proportion suggests that the Stay-Still condition had a greater increase in dwell proportion in correct versus incorrect trials compared to other conditions in the MCI- group, but not the MCI+ group. In other words, MCI- participants had an even greater increase in moved object dwell when correct in Stay-Still trials compared to other conditions, whereas MCI+ participants did not have a difference between conditions. This effect was on the threshold for

statistical significance, with the lower 95% confidence level just below zero ($\beta(5, 80) = 0.13,\ p = .051,\ 95\%\ CI\ [-0.001,\ 0.264],\ \eta_g^2 = .12$). Post hoc comparisons in Figure 4.14b shows that MCI- participants had a significant difference between the Stay-Still condition compared to other conditions in correct trials, whereas MCI+ participants did not.

Figure 4.14c shows mean dwell proportion on the moved object per participant against total percentage of correct trials, showing significant associations in MCI groups only. The weak and non-significant associations between these two measures in healthy groups seems at odds with correctness-split results from Figure 4.14a. This is likely the consequence of aggregating moved object dwell proportions to participant-level mean scores, which also combines conditions.

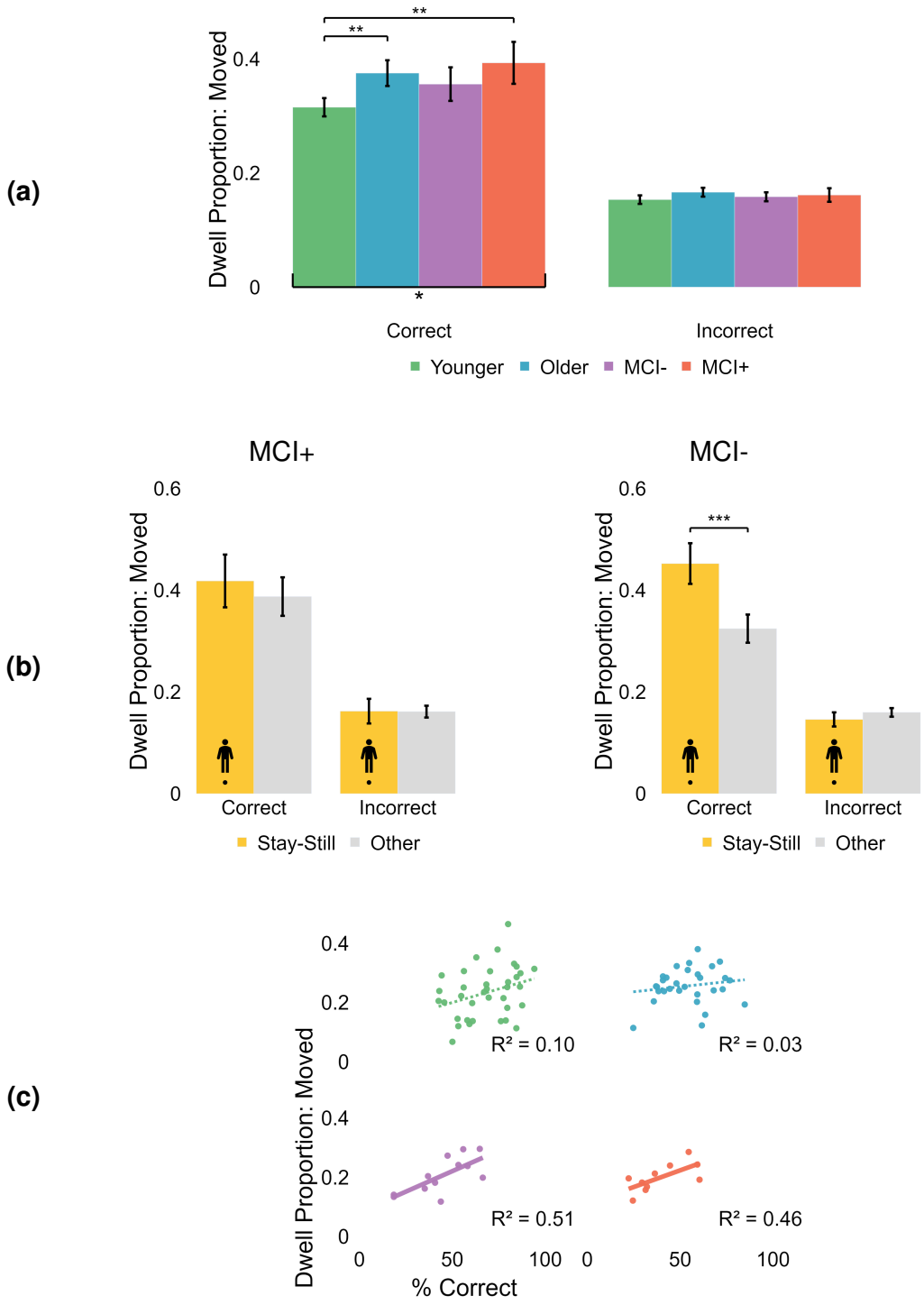


Figure 4.14. Relationships between task performance and dwell proportion on moved object at retrieval: (a) by group and trial correctness. Within-correctness significance is shown, between-correctness significance is not shown. (b) moved object dwell in Stay-Still trials compared to other conditions combined, split by correctness for MCI+ (left) and all other groups combined (right). (c) Scatter plots of mean dwell proportion against percent of correct trials, split by group. See Figure 4.7 for more details of scatter plots.

Dwell proportion on the stationary objects. There was a strong main effect of trial correctness on dwell proportion on the stationary objects ($F(1, 85) = 262.6, p < .001, \eta_p^2 = .76$), with Figure 4.14a visualising less dwell proportion on the stationary objects in correct trials compared to incorrect trials (significance not marked on the figure).

For healthy Younger and Older groups, there was a significant interaction between correctness and group on stationary object dwell ($F(1, 65) = 12.1, p < .001, \eta_p^2 = .16$), suggesting that the effect of age group on stationary object dwell varied by trial correctness. Figure 4.15a shows dwell proportion on the stationary objects split by trial correctness, with significantly less stationary object dwell in Younger versus Older participants for incorrect trials only. This means that, whereas correct trials drove the difference between healthy age groups for dwell proportion on the moved object, *incorrect* trials may have driven the difference between age groups in dwell proportion on the stationary objects.

The effect of correctness on stationary object dwell also varied by condition in healthy groups ($F(5, 195) = 2.87, p = .016, \eta_g^2 = .015$), suggesting that the difference in stationary object dwell for incorrect trials versus correct trials varied across conditions. Figure 4.15b (right) shows that this was due to less stationary object dwell for *correct* trials between the Stay-Still condition and other conditions, whereas there was no difference in stationary object dwell between conditions in *incorrect* trials. Note that there was no evidence that this effect varied between healthy Younger and Older groups as there was no three-way interaction between group, correctness and condition ($\beta(5, 195) = -0.016,\ p = .75,\ 95\%\, CI\ [-0.12,\ 0.08],\ \eta_g^2 = -.017$).

For MCI biomarker groups, there was no significant interaction between correctness and group on stationary object dwell ($F(1, 20) = 0.36, p = .55, \eta_p^2 = .018$), suggesting that the effect of biomarker status on stationary object dwell did not vary by trial correctness. This means that there was a similar difference in stationary dwell between MCI biomarker groups for correct and incorrect tri-

als. Figure 4.15a shows that the difference between MCI+ and MCI- groups is actually significant for incorrect trials but not correct trials based on post hoc pairwise tests. Still, the lack of interaction effect suggests that the non-significant difference between groups in correct trials was a similar difference.

There was a small significant interaction between trial correctness and condition in MCI groups $(F(5, 80) = 2.76, p = .023, \eta_g^2 = .032)$, suggesting that the effect of condition on stationary object dwell proportion varied by correctness in these participants. Similar to moved object dwell, Figure 4.15b (left) shows that MCI+ participants had no difference in stationary dwell proportion between Stay-Still and other conditions regardless of trial correctness. MCI- participants *did* show a difference here similar to Younger and Older groups; all three of these groups have been combined in 4.15b (right). Note that, despite this difference between groups, a three-way interaction was not significant when including all conditions, MCI groups only, and trial correctness $(\beta(5, 80) = -0.11,\ p = .12,\ 95\%\, CI\ [-0.25,\ 0.031],\ \eta_g^2 = -0.12)$.

Figure 4.14c shows mean dwell proportion on the stationary objects per participant against percentage of correct trials, showing no significant associations in any group. Again, the effect of aggregating dwell proportions per participant likely explains the discrepancy between the lack of significant correlations here and the main effect of correctness on stationary dwell.

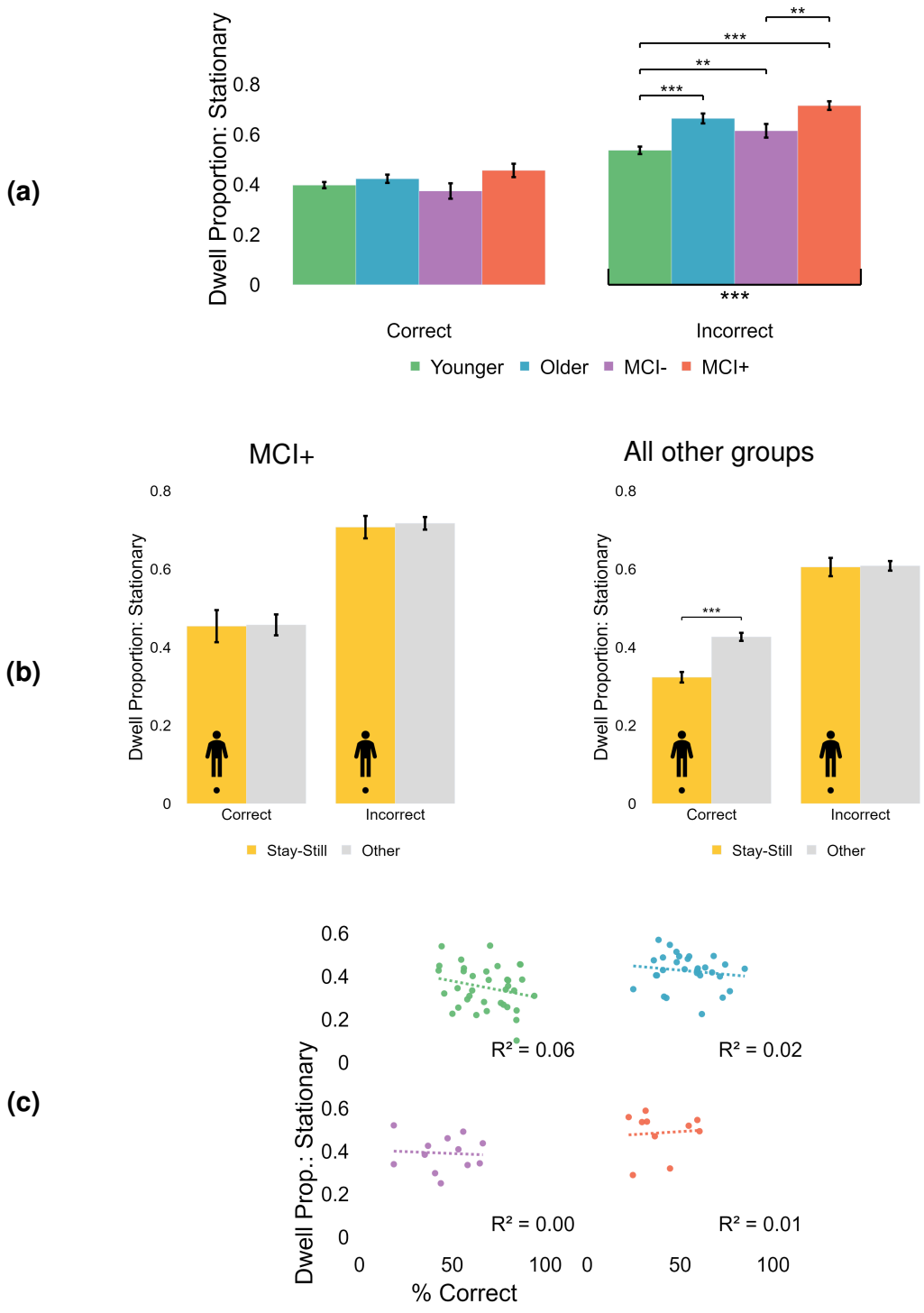


Figure 4.15. Relationships between task performance and dwell proportion on stationary objects at retrieval: (a) by group and trial correctness. Within-correctness significance is shown, between-correctness significance is not shown. (b) stationary object dwell in Stay-Still trials compared to other conditions combined, split by correctness for Younger, Older and MCI- groups combined (left) and MCI+ participants only (right). (c) Scatter plots of dwell proportion on the stationary objects against percent of correct trials, split by group. See Figure 4.7 for more details of scatter plots.

Dwell proportion on the table at retrieval. There was a small significant main effect of trial correctness on dwell proportion on the table at retrieval $(F(1, 85) = 7.87, p = .006, \eta_p^2 = .084)$. However, Figure 4.16a suggests that this is explained by reduced table dwell for incorrect versus correct trials for all groups except the Younger group, which showed the opposite effect (significance not marked on the figure). Indeed, there was a small but significant interaction between correctness and group when isolating Younger and Older groups $(F(1, 65) = 5.19, p = .030, \eta_p^2 = .074)$ but not MCI groups $(F(1, 20) = 0.49, p = .49, \eta_p^2 = .024)$. These results suggest that table dwell was more associated with incorrect trials in Younger participants, whereas it was more associated with correct trials in Older, MCI- and MCI+ participants. This effect in Younger participants is unexpected, as table viewing was hypothesised to be associated with performance in the positive direction. However, it is important to note that Younger viewing of the table was higher than other groups in both correct and incorrect trials, and this group also outperformed other groups in task performance (Section 4.4.1).

The effect of correctness on table dwell also varied by condition in healthy groups with a very small effect size $(F(5, 195) = 2.64, p = .024, \eta_g^2 = .008)$, suggesting that the difference in table dwell for incorrect trials versus correct trials varied slightly across conditions. However, a stronger three-way interaction shows that this effect varied by group $(\beta(5, 195) = -0.10,\ p = .005,\ 95\%\ CI\ [-0.031,\ -0.180],\ \eta_g^2 = -0.12)$, which can be explained by decreased table dwell in Stay-Still compared to other conditions in the correct trials in the Younger group, the opposite effect in the incorrect trials, and potentially the reverse of *both* of these differences for Older participants (Figure 4.16b). In other words, the Younger participants showed *increased* table dwell when *incorrect* compared to correct, with *more* table dwell during the Stay-Still condition compared to others when *incorrect* but *less* when *correct*. By contrast, the Older participants showed *less* table dwell when incorrect, with *less* table dwell

in the Stay-Still condition than other conditions when *incorrect*, but *more* when *correct*. However, note that these effects in Older participants were not significant, and may be better described as having a similar amount of table dwell across conditions.

For MCI groups, as mentioned there was no significant interaction between correctness and group on table dwell (see above), suggesting that the effect of biomarker status on stationary object dwell did not vary by trial correctness. This means that there was a similar difference in table dwell between MCI+ and MCI- groups for correct and incorrect trials. Figure 4.15a shows that MCI+ participants had significantly less table dwell than MCI- participants in both correct and incorrect trials from post hoc pairwise comparisons.

There was no significant interaction between trial correctness and condition in MCI groups $(F(5, 80) = 0.54, p = .74, \eta_g^2 = .004)$, suggesting that the effect of condition on table dwell proportion did not vary by trial correctness in these participants.

Figure 4.16c shows mean dwell proportion on the table per participant against percentage of correct trials, showing a significant associations for the Older group only. In Younger participants, the effect of trial correctness on table dwell was possibly masked here because the effect depends on the condition.

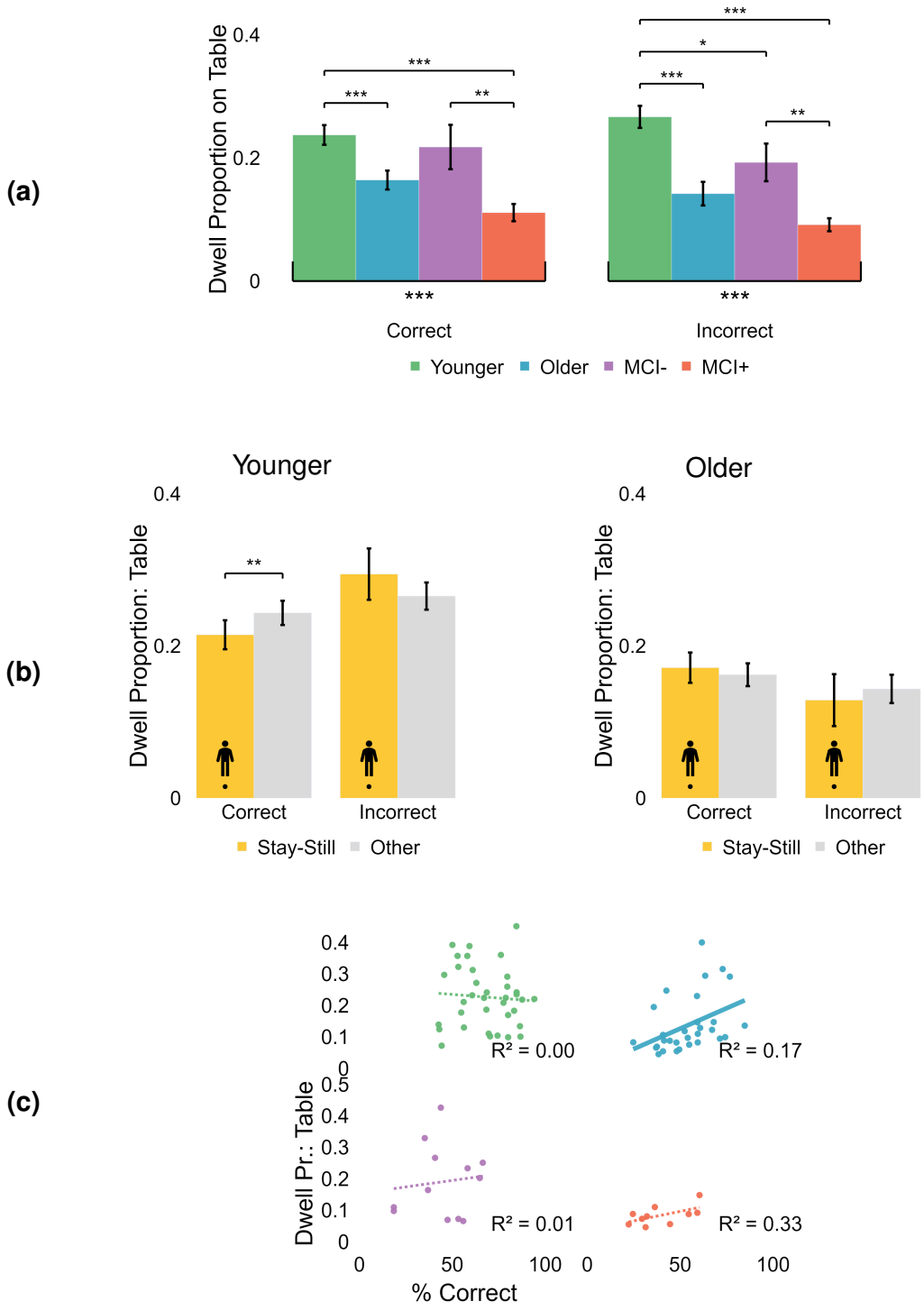


Figure 4.16. Relationships between task performance and dwell proportion on table at retrieval: (a) by group and trial correctness. Within-correctness significance is shown, between-correctness significance is not shown. (b) table dwell in Stay-Still trials compared to other conditions combined, split by correctness for Younger (left) and Older (right) groups. (c) Scatter plots of dwell proportion on the table against percent of correct trials, split by group. See Figure 4.7 for more details of scatter plots.

Encoding measures. No significant main effects of trial correctness were found on table dwell ($F(1, 85) = 0.16, p = .69$), object dwell ($F(1, 85) = 0.05, p = .81$), or scan-path entropy ($F(1, 85) = 1.02, p = .32$) at encoding.

There was no significant main effect of trial correctness on MultiMatch shape similarity between encoding and retrieval when including all groups ($F(1, 85) = 3.47, p = .07$), or just Younger and Older groups ($F(1, 65) = 0.005, p = .94$). However, there was a moderate main effect of trial correctness when including only MCI groups ($F(1, 20) = 20.1, p < .001, \eta_g^2 = .50$). For MCI groups, there was no interaction between group and correctness ($F(1, 20) = 0.79, p = .39$) or condition and correctness ($F(2, 32) = 1.55, p = .23$); but there was a weak yet significant three-way interaction between MCI group, condition and correctness ($\beta(2, 32) = -0.025,\ p = .044,\ 95\%CI[-0.0004, -0.051],\ \eta_g^2 = -0.026$). These results can be explained by the reduced MultiMatch shape similarity for MCI+ participants in *correct* Walk-Still trials as shown in Figure 4.17b. This suggests that only MCI+ participants had less similarity between encoding and retrieval phases, but only in correct Walk-Still trials.

Figure 4.17c shows mean MultiMatch shape similarity scores per participant against percentage of correct trials, showing a strong significant association for the MCI+ group only. Post-hoc analysis revealed that this effect was driven by the Stay-Still condition, which had a very strong positive Pearson's correlation in the MCI+ group ($r(8) = 0.87, p = .001$) compared to the Walk-Still and Teleport-Still conditions which both had non-significant correlations ($r(8) = 0.32,\ p = .36;\ \ r(8) = 0.42,\ p = .15$).

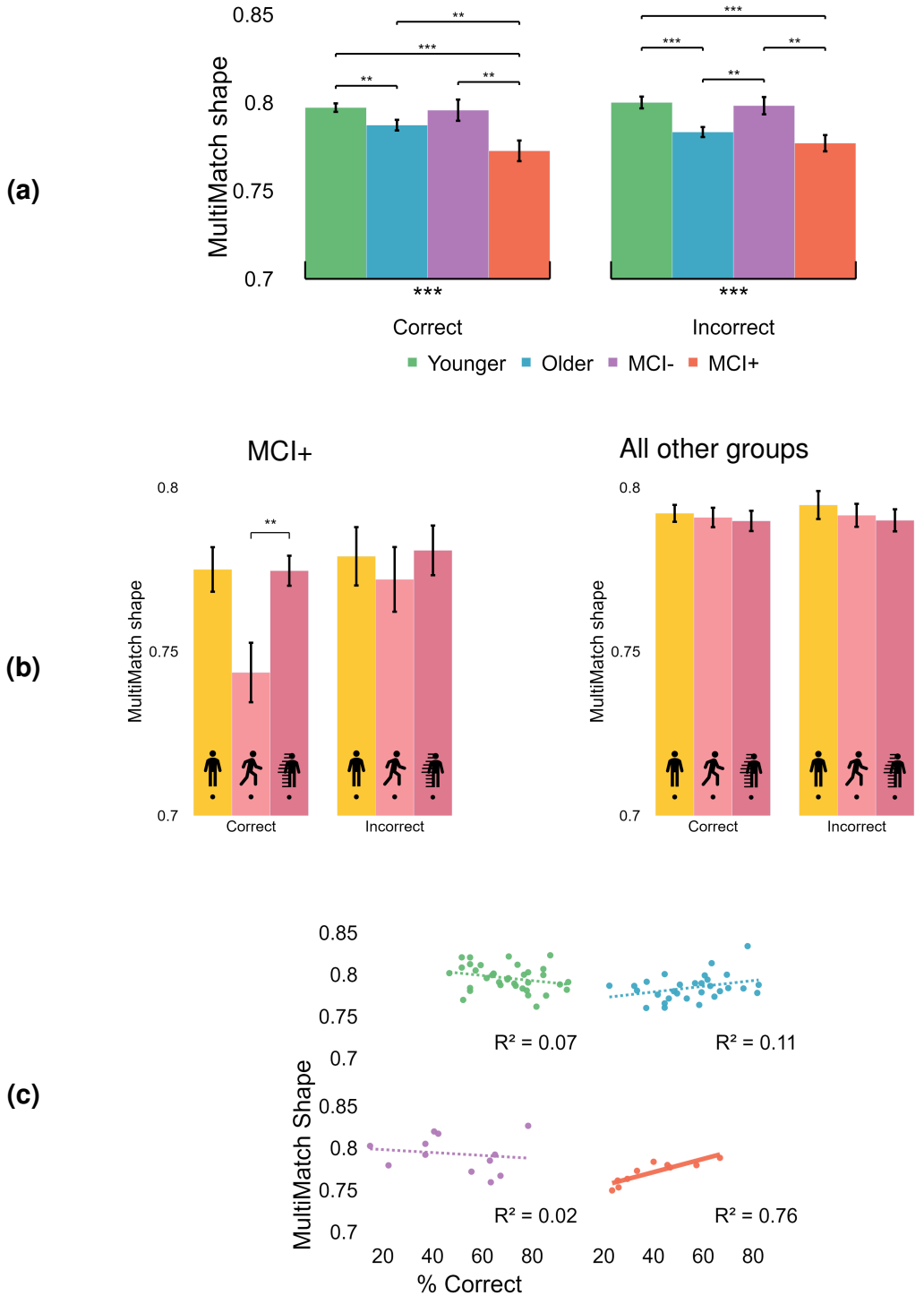


Figure 4.17. Relationships between task performance and Multimatch 'shape' similarity between encoding and retrieval scan-paths: (a) by group and trial correctness. Within-correctness significance is shown, between-correctness significance is not shown. (b) MultiMatch shape similarity in Stay-Still, Walk-Still and Teleport-Still trials, split by correctness for MCI+ participants (left) and Younger, Older and MCI- groups combined (right). (c) Scatter plots of MultiMatch shape against percent of correct trials, split by group. See Figure 4.7 for more details of scatter plots.

4.4.3.4 Post-hoc analyses: relative dwell measures at retrieval

Summary of key findings.

- Subtracting stationary object dwell time from table and moved object dwell time combined created a 'stationary avoidance' measure at retrieval, which showed large differences between MCI+ participants and all other groups in the Stay-Still condition.

Although not directly hypothesised, a natural follow-up to AOI-specific fixation patterns was to examine fixation *differences* between AOIs. Caution must be taken with interpretation of results here because there were numerous combinations of AOIs that could have been used to create difference measures. However, care was taken to only follow-up with post hoc comparisons that could be justified on the same grounds as hypotheses.

Stationary avoidance. A direct combination of the three hypothesised fixation areas of interest (AOIs) was calculated to create a relative measure of gaze patterns during retrieval. Although several combinations were available, the most logical and consistent with hypotheses involved subtracting AOI fixation time expected to *negatively* relate to performance from AOIs expected to *positively* relate to perfomance. Accordingly, stationary object dwell proportion was subtracted from the sum of moved object and table dwell proportions. I called the resulting measure 'stationary avoidance', because greater scores indicated less stationary viewing.

Figure 4.18 shows the results of this measure across groups and conditions. Again, the Stay-Still condition appeared to be the most discriminating for MCI+ participants ($MCI+ \ vs \ Older$: $t(38) = -4.10, p < .001, d = -1.50$; $MCI+ \ vs \ MCI\text{-}$: $t(20) = -4.68, \ p < .001, \ d = -2.00$), with significantly lower station-

ary object avoidance than the other three groups. Although, MCI+ participants had significantly lower stationary avoidance compared to all other groups in all conditions except Teleport-Still and Walk-Rotate conditions.

The Younger group scored consistently higher across conditions than other groups. Note that there was no interaction between group and condition for healthy groups on this measure ($F(5,\ 325) = 0.9,\ p = .47$), indicating a similar increase in Younger participants across all conditions. A significant group-condition interaction was found for MCI groups ($F(5,\ 100) = 3.4,\ p = .008$): the difference between MCI+ and MCI- groups in stationary avoidance varied across conditions.

This measure has the added advantage of showing mean positive bars for all but the MCI+ group in the Stay-Still condition, indicating not just a significant difference in this measure, but a difference in directional bias. Indeed, a positive bar indicates that participants viewed the table and moved object more than the three stationary objects combined.

The relationship between stationary avoidance and task performance was also examined, with findings corroborating results from moved object, stationary object, and table associations with trial correctness. Full results can be found in Appendix Section B.2.7.

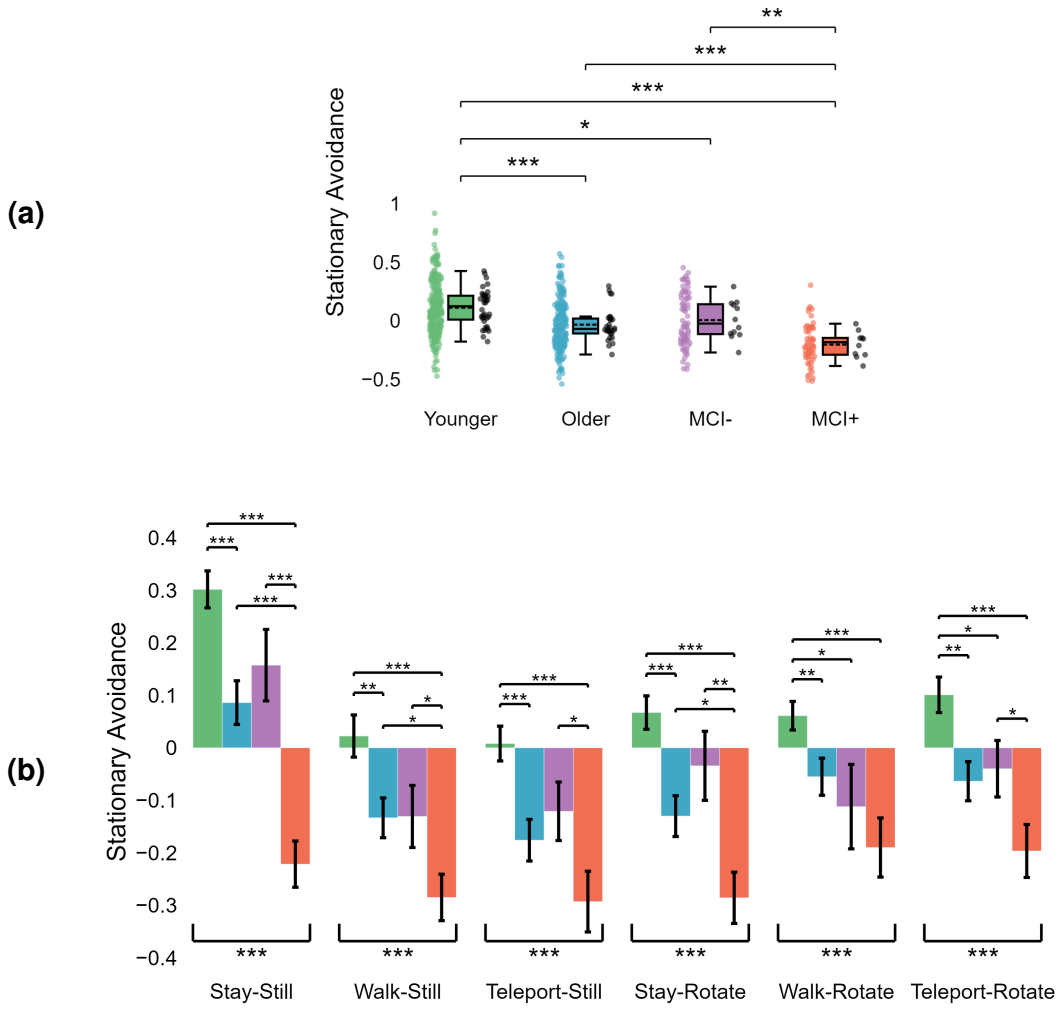


Figure 4.18. Dwell proportion difference on the moved object and table vs stationary objects at retrieval a.k.a. stationary avoidance. (a) Across all conditions between groups. (b) Within-condition group comparisons. See Figure 4.7 for more details.

4.5 Summary of chapter results

Results from this chapter are summarised here without wider discussion, which will be provided after all results chapters (in Chapter 7). The current section will be split into findings relevant to AD and healthy ageing, the latter focusing on differences between Younger and Older participants, with the former focusing on the main differences found for the MCI+ group compared to MCI- and Older participants.

4.5.1 AD-relevant findings

This chapter has presented results of behavioural and eye movement measures in relation to hypotheses that participants with early signs of AD will show memory impairment on the new VR task, as measured via spatial memory performance, memory monitoring, and eye movements. These results will be summarised by the main hypotheses and findings in this chapter.

The Walk-Still condition was *not* as discriminating as hypothesised. The Walk-Still condition was hypothesised to show the greatest difference between MCI+ participants and MCI- or Older participants. This condition was designed to require allocentric processing with availability of self-motion spatial updating. The Teleport-Still condition was also expected to be challenging for MCI+ participants because of the allocentric consistency of the object positions with encoding, similar to Walk-Still. All other conditions were expected to show similar performance between MCI+, MCI- and Older participants.

Although MCI+ participants seemed to perform lowest on the Walk-Still condition compared to other conditions, and the Older participants scored significantly higher in this condition, no difference was found between MCI+ and MCI- groups in this or any other condition on percentage of correct trials. There was also similar performance between the Walk-Still condition and the Teleport-Still condition in all groups for memory scores, confidence and eye movements, which was only expected for the MCI+ group due to impaired self-motion pro-

cessing. The pattern of results here suggests that self-motion in the Walk-Still condition was not enough to provide a memory advantage in healthy younger, healthy older or MCI- participants either.

MCI+ participants did have some different eye movements than other groups in the Walk-Still condition. For example, the dwell proportion on the stationary objects, and the 'stationary avoidance' measure separated MCI+ from other groups in this condition. However, this was not the largest difference between these groups for these measures, which was instead found in the Stay-Still condition.

There was one eye movement measure that showed a specific difference in MCI+ participants in the Walk-Still condition: the MultiMatch shape measure showed less similarity between encoding and retrieval scan-paths than other groups in this condition only. This supports the prediction that MCI+ participants have reduced gaze reinstatement effects due to damage in the medial temporal lobe. However, the same effect was absent for Teleport-Still trials, suggesting this may not have been driven by allocentric impairment. Furthermore, the lower MultiMatch shape similarity seemed to be specific to *correct* trials, suggesting that only when MCI+ participants selected the correct object did they have dissimilar scan-paths at retrieval compared to encoding, and only for the Walk-Still condition.

Taken together, some eye movement differences were observed between MCI+ participants and control groups, but results from Walk-Still trials do *not* support this condition as being strongly discriminating for these participants.

Replication of previous findings was mixed. Across groups, there was also no difference in performance or eye movement measures in the Stay-Rotate condition compared to the Walk-Still (or Teleport-Still) condition. This is different from previous results from similar paradigms (Simons and Wang, 1998; Wang and Simons, 1999), which provided evidence for an advantage of self-motion

to memory after viewpoint shifting (i.e. Walk-Still) compared to table rotation (i.e. Stay-Rotate). This suggests that the impact of shifting the view of the objects was the dominant effect on memory, with any differences afforded by self-motion being too subtle to detect with the power available. This may be due to the greater angular viewpoint shift in this study, which was 135° (compared to 40-47° in Wang and Simons' original publications), chosen to maximise reliance on allocentric processing.

A further caveat for the comparison between Walk-Still and Stay-Rotate was that the effect of walking on memory was not controlled for in Stay trials, unlike previous studies. Adding in a walking component to Stay conditions was avoided to allow comparison to Teleport conditions, with the consequence of reducing the comparability of these conditions. This may explain why group-level Stay-Rotate performance was sometimes higher than Walk-Still trials.

Performance on Move-Rotate conditions (i.e. with availability of egocentric processing) was generally very similar to Stay-Rotate, Walk-Still and Teleport-Still trials (i.e. conditions without availability of egocentric strategies) across groups. This is inconsistent with previous experimental results (Simons and Wang, 1998; Wang and Simons, 1999), at least for Stay-Rotate and Teleport-Still conditions, which were expected to lead to lower performance due to more spatial inconsistencies. Moreover, it does not support the explanation that the greater angular shift for non-egocentric trials explains the similar performance in viewpoint-shifting conditions (Stay-Rotate, Walk-Still and Teleport-Still conditions) because this should not matter for egocentric trials. Unless, that is, the amount of movement around the table was so disorienting (or increases time between viewings) as to make egocentric trials more difficult than if the angular shift was smaller.

Still, there was some evidence of improved memory in Move-Rotate (egocentric) conditions compared to non-egocentric conditions based on task performance, confidence and eye movements. This suggests that availability of

egocentric strategies did provide some memory advantage, supported by the significant effect of egocentric visual consistency of the object positions on several measures from factorial analyses.

An unexpected finding was that the MCI+ participants showed evidence of actually performing *better* than MCI- participants in the Walk-Rotate condition. Indeed, their task performance was comparable to Stay-Still trials in this condition, showing a specific effect of egocentric visual consistency, without an effect of self-motion consistency. This supports the hypothesis that MCI+ participants would be relatively impaired in their self-motion processing but not their egocentric spatial representations. It is unclear why this effect was not seen with Teleport-Rotate trials as well. However, factorial analysis suggested that there was a general performance-reducing effect of teleporting in healthy adults, which may exist in MCI participants (despite a null finding) but was masked by low power.

A floor effect may explain performance in MCI- participants, who seemed to perform similarly in all but the Stay-Still condition, without a performance boost from egocentric availability in other conditions (Walk-Rotate and Teleport-Rotate). This pattern of results suggests an effect of 'change' for these participants, reducing performance in any trial requiring any table or subject movement between encoding and retrieval. This may have masked any effects of spatial consistencies, obscuring differences due to spatial cognition for this group. However, it is noteworthy that these participants scored significantly greater than chance in all conditions.

In sum, results suggest that there may, as hypothesised, have been a self-motion impairment in MCI+ participants. However, the task was not able to discriminate between MCI biomarker groups based on this, potentially due to compensatory mechanisms supporting performance in MCI+ participants, or floor effects masking differences between groups. A floor effect may also have obscured any differences in allocentric impairment between these groups, as

performance in Walk-Still and Teleport-Still was comparable across all putative memory-related measures.

Eye movements in the Stay-Still condition were most discriminating. Unexpectedly, the most consistently discriminating condition between groups and conditions was the Stay-Still condition. This condition showed the greatest difference in task performance between Older and MCI+ participants, although this did not discriminate MCI+ from MCI- groups. Still, it is notable that MCI+ was the only group to have comparable task performance in the Stay-Still condition compared to other conditions.

Fixation time on the stationary objects and the table showed the greatest difference between MCI+ and MCI- participants in the Stay-Still condition, compared to other conditions and measures. This is despite an absence of significant differences between these groups in fixations on the moved object, similar to task performance. However, moved object fixation time did show a strong association with task performance across all groups, suggesting that there was a memory effect on this eye movement measure, as expected. However, fixation time on the stationary objects and the table showed greater differences between MCI+ and other groups than moved object viewing.

The post hoc eye movement measure that provided the strongest difference between MCI+ and other groups was the 'stationary avoidance' measure, a relative proportion between moved, table and stationary object viewing. Specifically, the difference between fixation times on the moved object and the table compared to the stationary objects discriminated the MCI+ group from all other groups in several conditions, but greatest in the Stay-Still condition. Although this method of measuring gaze behaviour was not directly hypothesised, it essentially combined hypothesised effects into one metric.

Differences were also observed in how MCI+ participants' eye movements interacted with task performance. Specifically, these participants did not show

a difference between conditions in either correct or incorrect trials. In contrast, MCI- and Older participants had greater fixation time on the moved object and less fixation time on the stationary objects for the Stay-Still condition compared to other conditions, but only when correct. This suggests that the differing fixation patterns in MCI+ participants may not be fully explained by the difficulty of the task across different conditions, because their eye behaviour responded differently to conditions than control groups when isolating correctness. One caveat to this finding is that the difficulty itself was not controlled, so there may have been a different number of correct trials between groups and conditions.

Eye movements at encoding did not show large differences between groups. So far, the summary of eye movement results has focused on viewing behaviours at the retrieval phase. Fixation times at the encoding phase provided weak discrimination between groups at best. This was predominantly measured by two dwell proportion metrics: dwell time on the table, and on all configuration objects. These were coarse measures of eye movements at encoding, essentially reducing fixation proportions down to one measure (because dwell time on the table was de facto the inverse of dwell time on the configuration objects). These results showed evidence of a slight *increase* in table viewing for MCI- participants compared to MCI+ and Older groups. Indeed, biomarker-negative MCI volunteers showed encoding fixation measurements more similar to Younger participants than any other group. Overall, these results do not suggest a strong effect of encoding fixation differences between groups for the purposes of AD diagnosis, contrary to hypotheses.

Similar to fixation times, scan-path entropy at encoding exhibited no differences between groups. Although entropy aimed to capture more eye movement information than dwell proportion measures, it may still be too coarse to detect group differences, or too heavily influenced by the configuration of objects. A discussion and further exploration of these issues can be found in Appendix A.

Memory monitoring may have been different for MCI groups. The final modality of measurement to summarise is confidence ratings, which did not show a clear difference between Older, MCI- or MCI+ groups across conditions, but did potentially support a hypothesised difference in memory monitoring abilities. Specifically, a positive association between task performance and mean confidence rating was found in healthy groups but not MCI groups. Furthermore, when quantifying the difference between participants' confidence ratings in correct and incorrect trials, MCI+ participants had significantly less difference than Older participants, suggesting that these participants' memory monitoring was less accurate. However, there was little evidence that this effect was specific to biomarker-positive MCI participants, especially since MCI- participants had the lowest correlation between confidence rating and task performance. This is contrary to hypotheses, where I expected confidence ratings to separate MCI biomarker status due to an AD-related impairment in memory-monitoring.

Conclusion of AD-related results and relation to neuropsychology. Overall, eye movements in the Stay-Still condition elicited the greatest differences between MCI+ participants and their control groups, whereas the Walk-Still condition—and more generally task performance—did not separate groups well.

Notably, the MCI groups did not significantly differ in performance on the 4 Mountains Test of allocentric memory either, despite significantly lower MCI+ scores on the Rey Delayed Recall measure. These results suggest that the samples differed in their memory impairments, but not based on allocentric spatial recognition tests. The implications of this finding will be explored in the general discussion in Chapter 7.

4.5.2 Healthy ageing effects.

A secondary aim of this study was to observe age-related effects in spatial memory and eye movements by comparing Younger and Older groups on the same measures as discussed above. And indeed, consistent differences were observed between these groups throughout behavioural and eye movement measures, suggesting that the task was effective at identifying age-related cognitive decline.

Younger participants scored significantly higher on the task than Older participants across all conditions. The same effect was observed for confidence rating, supporting hypotheses of reduced memory performance with age but intact memory monitoring.

Similar between-condition task performance was observed in both healthy age groups, suggesting that there was no specific impairment in allocentric memory conditions, as hypothesised. There was a greater difference in performance in Walk-Still and Teleport-Still conditions between groups, but this was not significant. As mentioned, a general reduction in performance and confidence in Teleport conditions may explain this: teleporting could have been particularly disorienting in this condition for older adults. However, this effect was too small to detect if it existed.

For eye movements, I did not find significantly greater dwell proportions on the moved object in younger participants compared to older volunteers—if anything, there was the opposite effect, with younger adults spending less time fixating on the moved object than older adults but only in correct trials. This is contradictory to the hypothesised decrease in preferential viewing with age. However, younger participants did fixate significantly less on stationary objects, as expected. This discrepancy compared to moved object viewing may be explained by the consistently greater dwell time on the table for younger participants. This effect was significant in all conditions and at both encoding and

retrieval viewings, suggesting a generally higher use of table-viewing strategies across the task in Younger participants. This supports hypotheses of increased table viewing with younger, healthier participants, potentially due to strategies involving the 'centroid' of the object locations, or the previous position of the moved object at retrieval.

However, there was evidence that table viewing was not associated with younger adults due to their greater performance. Indeed, Younger participants seemed to have more table viewing in *incorrect* trials, whereas Older adults showed the opposite. This suggests that table viewing at retrieval in Younger participants may have served a different purpose than for Older counterparts. Furthermore, this effect seemed to vary by condition, with younger adults' table viewing in the Stay-Still condition reduced compared to other conditions in correct trials but not in incorrect trials. This may point to different use of table viewing strategies depending on the spatial manipulation of the objects and viewpoints, with the no-change Stay-Still condition requiring less table viewing in younger adults. Older adults, on the other hand, had reduced table viewing across conditions and trial correctness, but a positive relationship between overall table viewing and task performance. This suggests less employment of table viewing strategies overall, but a boost to performance when used.

A discussion of these and other results can be found in Chapter 7, which provides a general discussion on all the findings in this thesis. Prior to this, two shorter chapters will introduce further analyses, focusing mainly on discrimination of MCI+ participants.

Chapter 5

Linear time dynamics of gaze across averaged viewing periods

Abstract

Group differences in gaze behaviour within viewing periods was investigated.

Methodology. For each participant, the proportion of trials spent looking at each area of interest (AOI) was calculated at each time-point of a normalised viewing period. At the group level, linear trends in AOI viewing behaviour were modelled using straight line fits, with statistical comparison between groups made by two methods: (1) permutation testing of bootstrapped group mean data, and (2) hierarchical mixed-effects modelling of participant-level data.

Hypotheses. This chapter was partially exploratory, but analyses were kept consistent with key hypotheses from Chapter 4 such as expecting group differences in gaze behaviour on the moved object, stationary objects, and the table.

Summary of results. Significant differences were found between MCI+ participants and other groups that were not detected from findings in Chapter 4. In the Stay-Still condition, MCI+ participants looked at the table less than other groups early in the retrieval phase, and increased their viewing of the moved object less than other groups. In the Walk-Still condition, MCI+ participants may have had greater stationary object viewing at the *end* of the viewing period than other participants, although this effect was greater in the Stay-Still condition. Additionally, MCI+ participants were less likely to look at the table at the *end* of the encoding phase than all other participants.

Limitations. Interpretation of findings should be made with caution due to the exploratory and post hoc nature of this analysis.

5.1 Introduction

Although some eye-tracking differences have already been detected between groups, the main eye movement measures introduced in the previous chapter were derived from a summary statistic per viewing, such as the dwell proportion on the moved object at retrieval. This approach loses potentially important information relative to the granularity in which it was collected. Indeed, eye-tracking data were collected up to 120 times per second, allowing us to examine gaze changes over time *within* each viewing period.

One way of examining the time dynamics of eye-tracking data is to plot dwell proportions over time across a normalised viewing period. Indeed, we can further examine viewing-level fixation measures by taking the foveated area of interest (AOI) at each time-point, average across trials per participant and then compare groups. This chapter will describe a methodology for achieving this, inspired by techniques from previous work (Oleson et al., 2017; Mirman et al., 2008; Maris and Oostenveld, 2007).

The analyses in this chapter were not planned before initiating data collection, and involved a degree of exploration prior to hypothesis-testing. The methodology was still developed for collecting the full dataset, but caution should be taken when interpreting findings, which are less scientifically powerful than results of prior predictions. However, as with the previous post hoc section, analyses were kept consistent with hypothesised measures. For example, gaze behaviour on the moved object, the stationary objects and the table were tested for time-dynamical effects. Like before, we would expect differences in MCI+ gaze behaviour in all of these measures, particularly for the Walk-Still condition. However, we have seen that the Stay-Still trials produced the most differences between groups, so a focus on this condition will also be presented.

5.2 Methods

Pre-processing. Each virtual object was categorised (e.g. moved object, stationary objects) to allow comparison across trials. Viewing periods were up-sampled to a common frequency (1000Hz), and clipped to 7000ms for comparability. Note that the task had a fixed viewing period of 7 seconds per viewing period anyway. Viewing periods with tracking loss towards the end were padded with substitute values, which were not used for the averaging process.

Calculating proportional viewing over time. For each condition per participant, pre-processed trials were stacked and averaged per AOI to output the proportion of trials with a gaze point on that object at each time-point. This equates to the probability that a participant was looking at, say, the moved object at each time-point for a particular condition's viewing phase. Each participant therefore had proportional viewing over time, which was averaged at each time-point within groups to get a distribution of proportional viewing per time-point per group. Figure 5.1 summarises this data aggregation process. When splitting data this way, viewing-level measures, such as dwell proportions, can be considered very similar to the area under the curve of these marginal probability time-series.

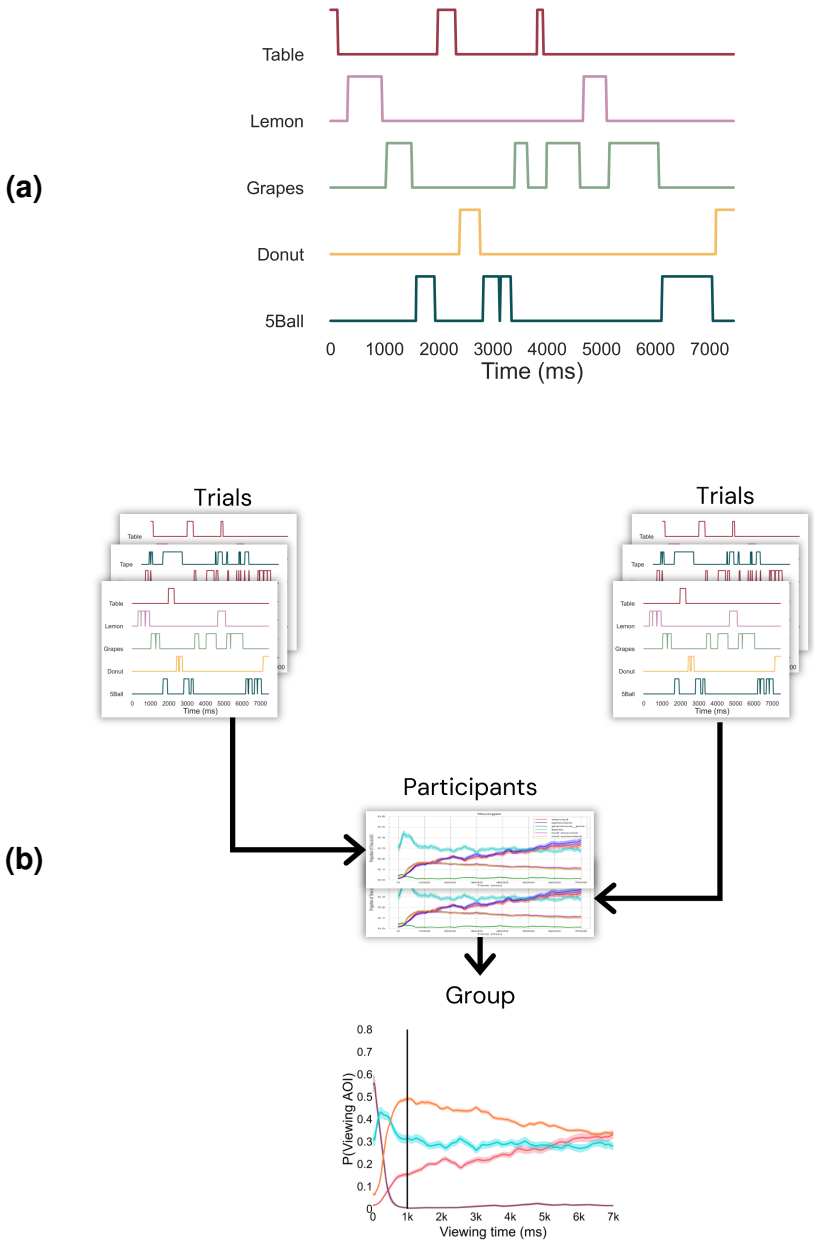


Figure 5.1. Demonstration of workflow to create proportional viewing time-points across an averaged viewing period. (a) Binary line plots showing fixations on different virtual objects in the environment. Gaze on an object can either be 1 or 0 at any time-point. Objects were categorised into AOI categories: moved object, stationary objects, table, and external (off the table). (b) Flowchart of averaging process from gaze data for each trial, which were then used for each participant to create a binomial distribution of proportional looking data at each time-point. These were averaged again at each time-point across participants to get group-level time dynamics.

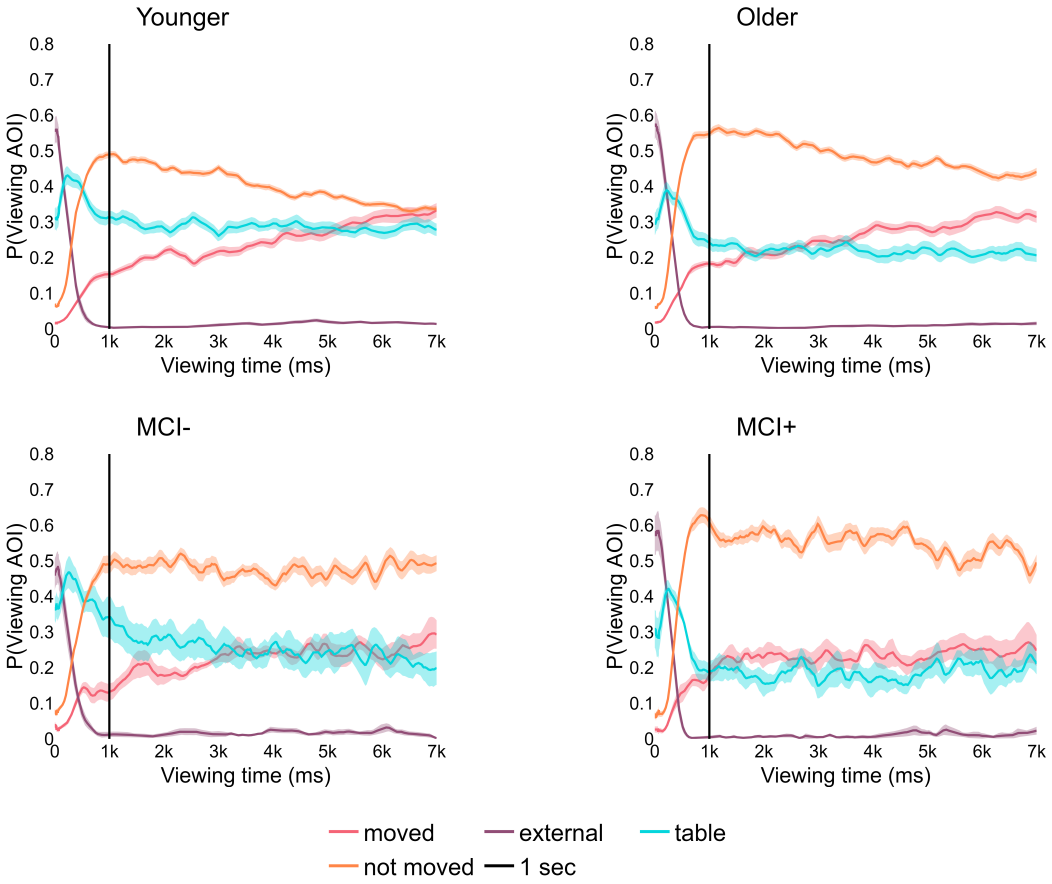


Figure 5.2. Proportional gaze on different AOIs over time for each group. These data are from retrieval-phase viewing periods, combining all conditions. Coloured bands are standard error zones around group proportions. The black vertical line is at 1 second (see main text).

Visual inspection of viewing over time revealed linear trends. Figure 5.2 shows averaged proportional looking on different AOIs over time for different participant groups. Visually inspecting these plots revealed similarities and differences in the amount and the change in viewing patterns for specific AOIs between groups. For instance, all groups had a spike and drop-off in table viewing within the first second, accompanied by increased viewing of table objects. This likely reflects the initial adjustment to the viewing period after the Occlusion Dome lifted up, which took 1s. After this first second, we can mostly observe linear trends over time for AOI viewing. For example, the Younger and Older

groups showed a clear up-trend in moved object viewing over the 7 seconds of the retrieval phase. Additionally, the proportion of looking at the table stayed at roughly 0.3 (30%) for the Younger group, unlike other groups, reflecting the increased proportion of dwell time on this AOI from viewing-level measurement (Figure 4.11). The MCI+ group had a particularly sharp drop-off in table viewing early on and a relatively flat line for the moved object compared to other groups.

5.2.1 Statistical analysis of group differences

To statistically test for differences in proportion looking over time, a procedure was developed that involved fitting straight lines to participant-level data to produce group-level linear trends in gaze behaviour. This produced a distribution of linear changes in proportional looking over time for each group. The procedure was based on the observation that from around 1 second, viewing patterns over time were linear. This assumption can be justified by the aforementioned task dynamics of the Occlusion Dome lifting up (which took exactly 1 second), after which visual stimuli were static. The effect of this can be visualised by the 'external' curves in Figure 5.2, which includes the Dome and all other objects off the table. Nevertheless, the interpretation of the following results should bear in mind the caveat that these were patterns observed from the data, not predefined procedures. A more complex procedure could fit non-linear curves to all proportional looking over time (as in Oleson et al., 2017), which would capture more time dynamics but lose the simplicity of the approach described here.

Another consideration was how to fit a line to time-series data which is highly auto-correlated. Oleson and colleagues addressed this by modelling residuals with a first-order autoregressor, albeit when fitting non-linear curves (Oleson et al., 2017). Indeed, fitting an Ordinary Least Squares (OLS) linear model to this type of data violates the assumption of unbiased errors for the classical linear model. This renders the variance-derived parameters of the model

unreliable for statistical significance. However, examining the residuals from participant data revealed consistent normal distributions with no heteroscedasticity around OLS fits (see Figure 5.3b for an example), suggesting that the model coefficients were a good fit to the data. Furthermore, variance-derived model parameters (such as standard error) were not used to determine statistical significance directly, as should become clear. Rather, significance was tested using permutation testing, as detailed in the next section.

However, linear model variance was used to account for within-participant variation through random re-sampling. This departs from Oleson and colleague's procedure, where bootstrap re-sampling (full data random sampling with replacement) was used on each participant's proportional viewing data and a random model was selected to account for within-participant variation. This approach was found to underestimate participant error when fitting OLS to the data. Instead, because the assumption of normality of residuals held (Figure 5.3b), I reasoned that the standard deviation of residuals could be used to repeatedly randomly sample participant linear trends by randomly generating two points based on the OLS fit and the standard deviation around it. This was a more conservative approach than Oleson and colleagues' procedure, where the data itself was bootstrap resampled to estimate error. Figure 5.3a visualises some of this new approach.

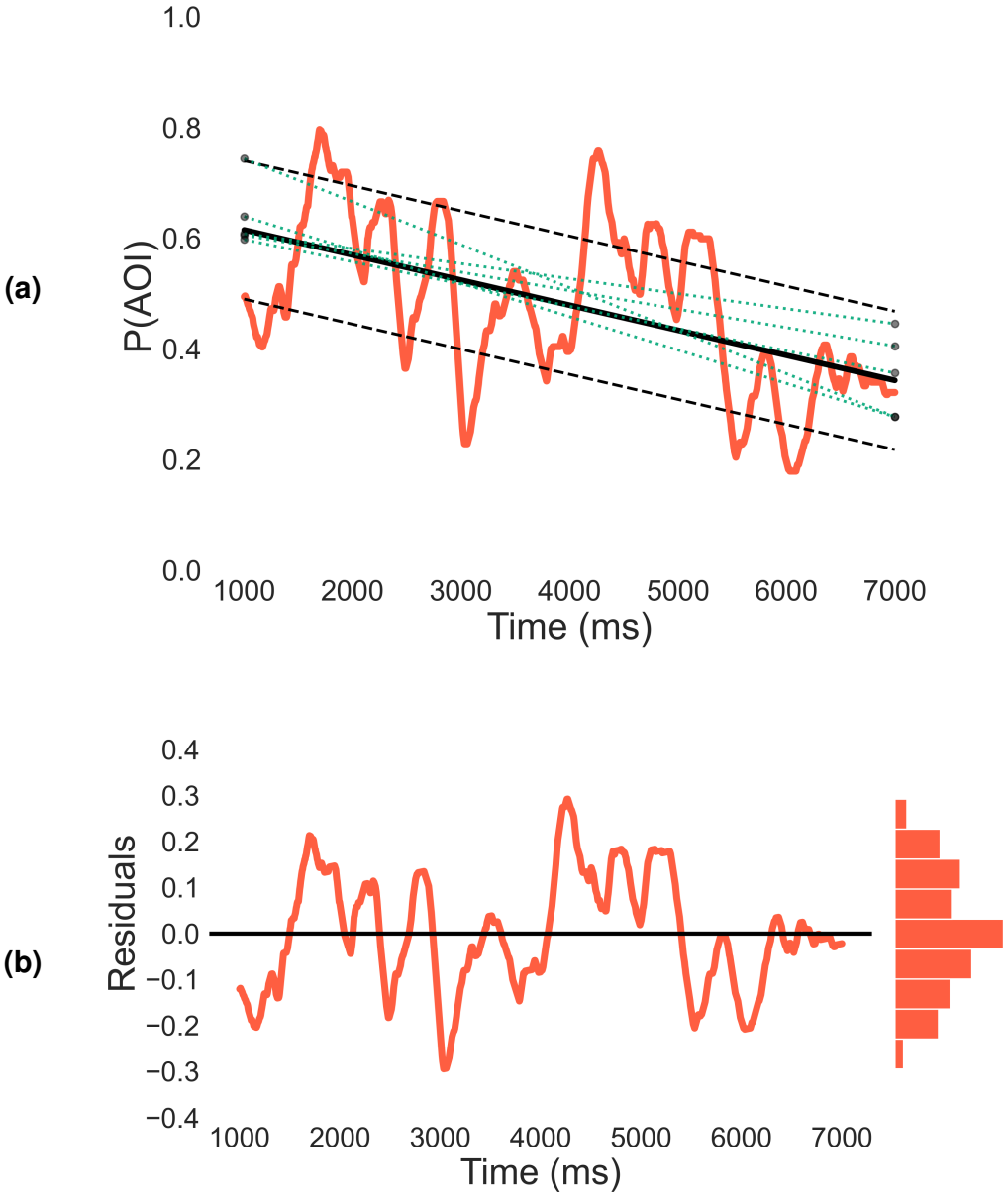


Figure 5.3. Examples of participant-level sampling from proportional looking data. (a) One participant's proportional looking data over time for a specific AOI (the example here is for stationary objects, but the method applies to all AOI categories). The red line is the probability of viewing that AOI over time for one participant. The solid black line is the OLS fit to estimate directional trend; the dashed black lines show the standard deviation of the residuals around the OLS fit; and the green dotted lines show five random within-participant samples used when estimating within-participant variation. Over enough repetitions, the average linear direction converges on the OLS fit. (b) The residuals around the OLS fit from (a), with the distribution of residuals on the right edge showing an approximately normal distribution around the OLS fit.

5.2.1.1 Method 1: comparing bootstrapped mean linear trends at the group level

Two methods were used to compare linear viewing trends. The first involved comparing bootstrapped mean linear models of the data, with permutation testing for significance (Maris and Oostenveld, 2007). The procedure went as follows for each AOI:

1. Fit an Ordinary Least Squares (OLS) linear model to each participant's proportional looking over time (such as in Figure 5.3a).

2. Save the standard deviation of residuals around the OLS line for each participant (dashed black lines in Figure 5.3a).

3. For the first and last time-points (i.e. 1000ms and 7000ms), randomly sample from a normal distribution for the y-axis value (proportion of trials looking at that AOI for that time-point), using the OLS intercept for that time-point as the mean and the standard deviation of the residuals as the standard deviation. Take the slope and y-intercept (where $y = 0$) of the line between these two points as the linear model for that participant. (green dotted lines in 5.3).

4. Using these randomly sampled linear trends per participant, bootstrap re-sample participants within each group (i.e. randomly sample with replacement).

5. Take the mean of the linear parameters (slope and y-intercept) as the linear trend for each group.

6. Repeat steps 1-5 many times to get a distribution of mean linear trends per group.

By bootstrap re-sampling within-participant lines first (steps 1-3) before re-sampling participants to estimate group mean lines (steps 4-6), more within-

participant variation can be accounted for than if the overall OLS fit per participant was taken (i.e. if steps 2 and 3 were skipped).

Measures for group comparison. This procedure was followed for several key proportional-time relationships: the table, the moved object, the stationary objects, and stationary avoidance (from Section 4.4.3.4) over time. This latter measure involved subtracting stationary object proportional looking from the sum of moved and table viewing at each time-point.

There were several metrics that could be derived from linear parameters to compare between groups. For instance, examining differences in slopes allowed for comparison of change in viewing behaviour over time. I also compared proportional looking at regular time-points along the whole time series. However, for simplicity, I focused mainly on time-specific differences at the beginning (1 second, or the difference in viewing behaviour early in the viewing period) and at the end (7 seconds, or the difference in viewing behaviour just before the selection phase begins).

These comparisons could not be made by looking at viewing-level summary measures alone. However, I also examined the area under the line to approximate total proportional looking per group, similar to dwell proportion measures.

To account for the large number of comparisons being made, permutation testing was employed to determine statistical significance. This first involved shuffling data into random groups and generating pairwise comparisons of slopes, one-second intercepts, seven-second intercepts, and areas under the lines over many repetitions $(R = 1000)$. Note that time-point comparisons at one- and seven-second y-intercepts were tested against the probability of any time-point with that effect or greater. Time-point comparisons were made every 500ms, starting at 1s and ending at 7s.

Pairwise comparisons from the real groups were compared to the distribution from shuffled data with one-tailed alpha levels $(\alpha = .05)$, the direction

of which was consistent with results from Chapter 4. Results from this section were reported based on statistical significance alone, whereby a significant result signified that less than 5% of permutations for that comparison showed an effect of that size or greater.

5.2.1.2 Method 2: Mixed-effects modelling of within-participant random re-sampling

The second approach was developed to model participant data within a nested hierarchical structure. This method simply followed steps 1-3 from the procedure in the previous section, with 1000 repetitions for each participant per analysis. For each pairwise comparison, I then ran a hierarchical mixed-effects ANOVA predicting group onto a linear parameter (e.g. slope of the line) with random effects of participant.

This method was employed to allow closer comparison between time-dynamical results and viewing-level analyses, which also made use of hierarchical mixed-effects models of within-participant data. It also provided a more conservative method of comparing group differences while maintaining the power of hierarchical data.

Results of significant pairwise mixed-effects models were reported with *z*-scores, *p*-values and Cohen's *d* effect sizes based on participant means.

5.3 Results

Figures 5.4 and 5.5 show group-level bootstrapped linear models of proportional looking on key AOIs over time for Stay-Still and Walk-Still conditions. Other conditions were not presented here: results presented in this section were chosen selectively based on key findings from hypothesised eye movement results. Comparisons that are not described here were less relevant to the discrimination of MCI+ participants from other groups and did not show differences between groups in viewing-level measures.

In general, statistical significance by the second method—involving hierarchical mixed-effects modelling of within-participant variation, hereafter referred to as 'method 2'—was always more conservative than the first method (permutation-based testing of bootstrapped group parameters, or 'method 1'). Accordingly, results of pairwise significance are marked on Figures 5.4 and 5.5 by two levels of significance: a comparison was either significant by method 1 alone, or by both methods 1 and 2, indicated using dashed and solid lines, respectively.

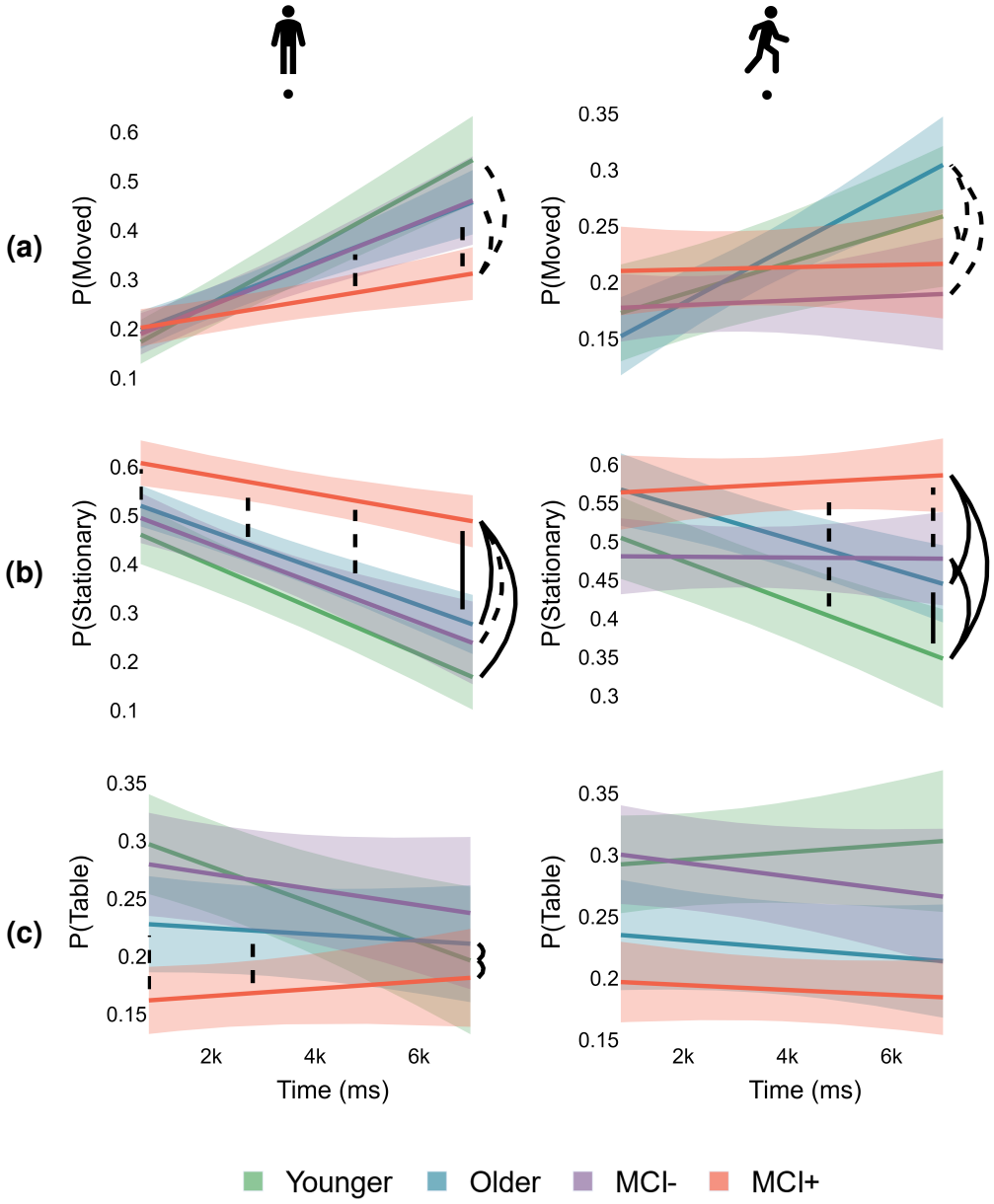


Figure 5.4. Bootstrapped linear time dynamics of proportional looking at different AOIs per group. The left-hand column are all from the Stay-Still condition; the right-hand column from the Walk-Still condition. Curved black lines represent significant differences between slopes of the lines; vertical black lines represent significant time-point differences between one group and all other groups, i.e. the most extreme group at that point is significantly different from all other groups. Dashed lines represent significance by the first approach (permutation testing); solid lines represent significance by both the first and second approach (hierarchical mixed-effects modelling of within-participant variation). All significance lines are $p < .05$ (no further p-value stratification was included). (a) The moved object. (b) The stationary objects. (c) The table.

Moved object. Figure 5.4a shows moved object linear time dynamics in the Stay-Still and Walk-Still conditions. The MCI+ group had a significantly lower slope coefficient than the other three groups in the Stay-Still condition based on method 1. The MCI+ group also had significantly lower proportional viewing of the moved object at the seventh second (7000ms) than the other three groups by method 1. These two results were not statistically significant when tested using method 2 $(p > .05)$. All groups started with similar proportional looking at the moved object in this condition, but the MCI+ group had less of an increase over time compared to the other three groups, resulting in a significantly flatter slope and lower likelihood of viewing the moved object by the end of the viewing period, just before the selection phase.

For the Walk-Still condition, MCI+ participants had significantly flatter slopes than both Older and Younger participants by method 1, but no difference was found between MCI+ and MCI- groups. MCI- participants also had significantly flatter slopes than Older participants when tested by method 1. No effects were significant when tested by method 2. These results suggest that MCI participants were less likely to increase viewing of the moved object over the course of Walk-Still retrieval phases when compared to healthy groups.

Stationary objects. Figure 5.4b shows bootstrapped linear models of stationary object viewing. In the Stay-Still condition, MCI+ participants had significantly higher proportional gaze at the seventh second than all other groups in both statistical approaches $(MCI+ \ vs \ MCI\text{-}: \ z(20) = 2.69, \ p = .007, \ d = 1.03; \ MCI+ \ vs \ Older: \ z(38) = 3.48, \ p = .001, \ d = 0.98; \ MCI+ \ vs \ Younger: \ z(45) = 5.58, \ p < .001, \ d = 1.56)$. Significant differences between MCI+ and other groups were also found at several time-points over the averaged viewing period under method 1. These results suggest that, although MCI+ participants were probably more likely to view the stationary objects at any point in a Stay-Still viewing period, this effect was most detectable just before the selection

phase.

MCI+ participants also had significantly flatter slopes than Older ($z(38) = 2.18,\ p = .029,\ d = 0.56$) and Younger ($z(45) = -2.41,\ p = .016,\ d = 0.66$) groups when tested using both statistical methods, but only by method 1 compared to MCI- participants. This indicates that MCI+ participants tended to decrease their viewing of the stationary objects less than other participants in the Stay-Still condition, although this effect may not discriminate MCI+ and MCI- participants well.

Not shown in Figure 5.4b (left), MCI+ participants had significantly greater areas under the lines compared to all other groups ($MCI+\ vs\ MCI\text{-}:\ z(20) = 4.03,\ p < .001,\ d = 1.36;\ MCI+\ vs\ Older:\ z(38) = 3.64,\ p < .001,\ d = 0.99;\ MCI+\ vs\ Younger:\ z(45) = 6.31,\ p < .001,\ d = 1.70$), indicating more overall viewing of the stationary objects in the Stay-Still condition, consistent with viewing-level dwell proportion on the stationary objects (see Section 4.4.3 in the previous chapter).

In the Walk-Still condition, both MCI groups had statistically greater slopes than Younger and Older groups when tested by method 2 ($MCI+\ vs\ Older:\ z(38) = 2.07,\ p = .038,\ d = 0.56;\ MCI+\ vs\ Younger:\ z(45) = 2.54,\ p = .011,\ d = 0.56$; $MCI\text{-}\ vs\ Older:\ z(40) = 2.19,\ p = .028,\ d = 0.62;\ MCI\text{-}\ vs\ Younger:\ z(47) = 2.68,\ p = .007,\ d = 0.62$). Inspection of Figure 5.4b (right) suggests that MCI participants had less of a decrease in stationary viewing than healthy participants, consistent with time dynamics of moved object viewing for these groups. MCI+ participants also showed significantly greater likelihood of viewing the stationary objects than all other groups at approximately 4.5 seconds and 7 seconds when tested by method 1, but not method 2.

In addition, the Younger group had significantly less likelihood of viewing the stationary objects than all other groups at the seventh second when tested by both methods 1 and 2 ($Younger\ vs\ Older:\ z(20) = -2.04,\ p = .041, d = -0.39;\ Younger\ vs\ MCI\text{-}:\ z(47) = -2.17,\ p = .030,\ d = -0.56;\ Younger\ vs\ MCI+:$

$z(45) = -3.40,\ p = .001,\ d = -0.94$). Younger participants were least likely to view the stationary objects just before the start of the selection phase.

Table. Figure 5.4c shows group bootstrapped linear time changes in table viewing in the Stay-Still and Walk-Still conditions. For Stay-Still trials, MCI+ participants were less likely to view the table at 1 second and 3.5 seconds by method 1, but not method 2. This suggests that MCI+ participants may have been less likely to look at the table early on in Stay-Still retrieval phases on average, but this effect was not large enough to detect with method 2.

Additionally, the Younger group had significantly more declining slopes than both MCI+ and Older participants according to methods 1 and 2 ($Younger\ vs\ Older:\ z(65) = -1.98,\ p = .048,\ d = 0.35;\ Younger\ vs\ MCI+:\ z(45) = -2.00,\ p = .046,\ d = 0.53$). On average, Younger participants appeared to reduce their viewing on the table over the course of Stay-Still retrieval phases than other participants, although this effect was moderate, with no significant difference compared to MCI- participants from either statistical method.

No significant slope or time-point differences were found between groups for the Walk-Still condition at retrieval for table viewing.

Table (encoding). Figure 5.5 (right) shows bootstrapped linear trends in table viewing for all trials at the encoding phase. MCI+ participants had a significantly lower likelihood of looking at the table at 7 seconds compared to all other groups ($MCI+\ vs\ MCI-:\ z(20) = -2.74,\ p = .006,\ d = -1.03;\ MCI+\ vs\ Older:\ z(38) = -2.14,\ p = .032,\ d = -0.69;\ MCI+\ vs\ Younger:\ z(45) = -2.64,\ p = .008,\ d = -0.89$). Inspection of the figure suggests that this was because of a combination of a lower likelihood of looking at the table from that start of the viewing period compared to MCI- participants (significant by method 1 only) and a decrease in this likelihood over time. Indeed, MCI+ participants had significantly more declining slopes than Younger and Older

participants ($MCI+\ vs\ Older:\ z(38) = -2.89,\ p = .004,\ d = -0.84;\ MCI+\ vs\ Younger:\ z(45) = -2.09,\ p = .036,\ d = -0.67$), who were more likely to increase their viewing of the table over the course of the encoding phase. These effects were not detectable by viewing-level analysis (Figure 4.12 in the previous chapter), but they suggest that MCI+ participants had different eye movement behaviour during memory encoding compared to control groups.

Stationary avoidance. Figure 5.5 (left) shows bootstrapped linear trends in stationary avoidance over time, calculated as the likelihood of looking at the stationary objects minus the table or moved object at each time-point. MCI+ participants generally scored lower across time-points than other groups, as well as significantly flatter slopes, when testing via method 1. However, none of these time-variant effects were significant by method 2.

Not shown in Figure 5.5, the MCI+ group had significantly smaller areas under the line compared to all other groups ($MCI+\ vs\ MCI\text{-}:\ z(20) = -4.59,\ p < .001,\ d = -1.71;\ MCI+\ vs\ Older:\ z(38) = -4.07,\ p < .001,\ d = -1.09;\ MCI+\ vs\ Younger:\ z(45) = -6.38,\ p < .001,\ d = -1.71$), consistent with viewing-level findings that, overall, MCI+ participants weighted their viewing behaviour towards stationary objects significantly more than other participants.

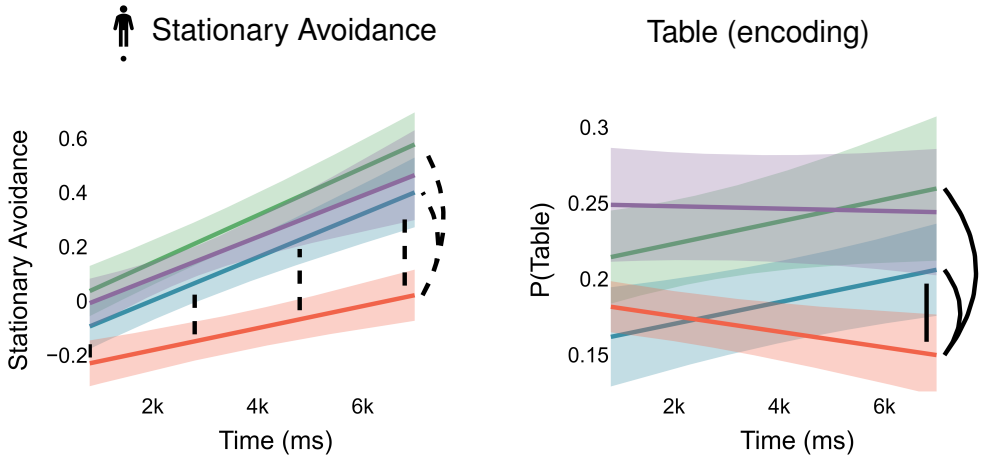


Figure 5.5. (Left) Bootstrapped linear trends of stationary avoidance in the Stay-Still condition at retrieval. (Right) Bootstrapped linear trends of table viewing across the encoding viewing period. An equal but opposite pattern of results exists for time dynamic viewing of configuration objects at encoding.

5.4 Summary and discussion

This chapter introduced a methodology for analysing the linear time dynamics of gaze behaviour across a normalised viewing period. By modelling changes in proportional gaze over time, I examined viewing behaviour that was hidden by more coarse-grained measures. Using two statistical approaches for group comparisons, I found that MCI+ participants had differences in their viewing behaviour that were not detectable by viewing-level results in Chapter 4.

Statistical significance. Some of these differences in viewing behaviour were only detectable by the less conservative statistical approach. This involved bootstrap re-sampling of group mean scores, with subsequent permutation testing. This methodology was not less conservative by design; several steps were taken to ensure rigour in statistical testing. For instance, a conservative approach was taken to within-subject variation than previous methods (Oleson et al., 2017). Moreover, permutation testing, while time-consuming, was a direct way of creating a comparative normal distribution for group comparisons based on random shuffles within the dataset itself. As such, a degree of confidence can be held in statistical significance derived from this method, which likely indicates that the mean scores between groups were not the same (i.e. rejecting the null hypothesis), at least for the sample in this study. Indeed, statistically significant findings from this first approach tended to corroborate differences in line parameters observable from visual inspection of Figures 5.4 and 5.5.

Nevertheless, a second method was included to model data at the participant level. This was originally developed to feed time-dynamical measures into the classification pipeline in the next chapter, which required one measure per independent observation (i.e. per participant). However, it also provided a more conservative approach than permutation testing for determining statistical significance in this chapter. Consequently, the strongest effects were highlighted

if they reached significance in both approaches. And again, viewing behaviour on the stationary objects showed the largest differences between MCI+ participants and MCI- or Older participants. This is consistent with findings from the previous chapter, showing that MCI+ participants' preference for viewing stationary objects tended to occur across the whole Stay-Still retrieval phases.

The Stay-Still condition was most discriminating again. As in Chapter 4, the Stay-Still condition was most discriminating between MCI- and MCI+ participants, with no differences between these groups in the Walk-Still condition by the more conservative approach. In fact, similar gaze behaviour was found between MCI groups in both moved and stationary object viewing, with relatively little change in likelihood of viewing these AOIs over time compared to healthy groups. This is contrary to hypotheses, which predicted greatest discrimination between biomarker groups in the Walk-Still condition. However, it is consistent with findings in previous sections, and provides stronger evidence that eye movements were similar in this condition between MCI groups, regardless of biomarker status.

There was evidence that the change in average viewing of the moved object over time was different for the MCI+ group compared to all other groups. Specifically, where other groups tended to increase the probability of viewing the moved object over time in Stay-Still retrieval phases, the MCI+ group showed less of an increase in this likelihood. Similarly, the likelihood of looking at the moved object just before the selection phase was lower for MCI+ participants than other groups: an effect that is masked when examining the whole viewing period (such as in Figure 4.9b in the previous chapter). These findings support hypotheses of reduced preferential viewing of the moved object in participants with early AD, suggesting that this effect only becomes detectable at finer-grained time periods for this task.

MCI+ participants had reduced table viewing at the *end* of encoding periods. Another notable finding was that MCI+ participants showed a reduced likelihood of looking at the table at the end of the encoding period compared to all other groups. Although the size of this effect was not as strong as some retrieval differences, it does support hypotheses that MCI+ participants would show reduced table-viewing behaviour during memorisation phases. Moreover, the data seem to indicate that MCI+ participants also decreased their viewing of the table over time across the encoding period, whereas healthy groups showed the opposite effect. The inverse of this (not shown in figures) was an increased likelihood of viewing the objects towards the end of the encoding period. Interestingly, evidence from symptomatic familial AD carriers suggests the opposite effect: that AD participants have *reduced* fixation time on the most salient visuospatial stimuli (Pavisic et al., 2021). A closer examination of what explains this discrepancy may reveal important differences in how participants with AD encode visuospatial information.

Caveats to interpretation. As mentioned, interpretations of this chapters' results should be made with the caveat that the analysis was developed partially post hoc. Particular care must be taken in confirming previous predictions that were null based on pre-planned measures. Indeed, there was evidence of a reduced preferential viewing effect on the moved object in MCI+ participants by this finer-grained methodology, but not by better-established viewing-level summary statistics presented in the previous chapter. Accordingly, effects that are novel for this analysis should be used as a foundation for future predictions, hypotheses or even task design for further exploration. For example, the linear models of gaze behaviour over time predict an even greater difference between groups if viewing periods were extended beyond the seven seconds used in this task. A future iteration of this paradigm could use the findings in the current work to support predictions of reduced preferential viewing effects given

extended viewing at retrieval.

Another important caveat to interpretation of results in this chapter is that several analyses have not been presented due to time and space constraints. Specifically, there were at least 3 measures per linear model (e.g. slope, 1 second intercept, 7 second intercept) each for at least 3 AOIs, with up to six pair-wise group comparisons. It was not feasible to include these for another four conditions, let alone examining between-condition comparisons. Of course, this means that a wealth of data remains for potential future exploration, but for the current work, targeted selection was required. Accordingly, AOIs and conditions most relevant to previous findings (i.e. the Stay-Still condition) and key AD-related hypotheses (the Walk-Still condition) were focused on. And indeed, the relevance of the main findings to hypotheses is a promising sign of wider significance.

Discussion of results: the effect of a forced viewing period. If we assume that some of the findings are valid, it is worth exploring why they may exist. One explanation may lie in a particular characteristic of the task: that there was a forced period of viewing prior to object selection (i.e. the retrieval phase). This design feature provided a fixed length of time to analyse gaze behaviour at retrieval, also allowing closer comparison to encoding periods. Analysing gaze data at retrieval has been avoided in previous similar studies *without* forced viewing because too many trials would have too little retrieval time to examine meaningful eye movement data (because some participants will select an object immediately; e.g. Coco et al., 2023; Segen et al., 2021; Hilton et al., 2020; Muffato et al., 2019). Incidentally, including this forced viewing precluded reaction time or decision-making analyses because the participant may have had to wait until the end of the viewing period before selecting the object they had already chosen. I observed anecdotally that, while waiting, some participants would fixate on the chosen object until they were able to select it. Participants

that chose quicker would have started this continuous fixation earlier. Conversely, some participants may not have decided on the object until the retrieval phase was over. Accordingly, the flatter slope for moved object viewing from MCI+ participants in the Stay-Still condition (Figure 5.4a [left]) may represent either lower memory performance (e.g. they fixated more on one of the stationary objects, which they thought had moved) or slower reaction times (e.g. they were still visually exploring which object may have moved). The latter explanation seems unlikely to fully explain the observed effect because MCI- participants showed comparable or even slower reaction times compared to MCI+ participants during neuropsychological testing (see TMT B and DST results from Figure 4.3). A memory effect seems more likely here based on two findings in the previous chapter: (1) the MCI+ group showed a mean reduction in task performance in this condition compared to other groups—albeit not reaching statistical significance—which eye movements may be a more powerful method to detect; and (2) MCI+ participants had a significant association between dwell time on the moved object and percentage of correct trials, suggesting that preferential viewing of the moved object is an indicator of memory performance.

Further exploration of results will be discussed in Chapter 7. For now, this chapter has demonstrated that a finer-grained analysis of gaze changes over time can provide potentially useful detection of AD-related gaze differences.

Chapter 6

Classification of biomarker-positive participants using task measures and conventional neuropsychological tests

Abstract

Task measures and traditional neuropsychological tests were compared for their diagnostic discrimination of MCI+ participants.

Methodology. A cross-validated supervised learning pipeline was used to compare single- and multi-feature logistic regression and support vector classifiers in their ability to discriminate MCI+ participants from (a) MCI- participants, and (b) both MCI- and Older participants combined.

Hypotheses. Significant differences found in previous chapters were expected to have the greatest discrimination by this methodology. In addition, a combination of task modalities (e.g. task performance plus eye movements) was expected to outperform univariate models. Importantly, task measures were expected to more accurately classify MCI+ participants than traditional neuropsychological tests.

Summary of results. Stationary avoidance in the Stay-Still condition showed the greatest univariate classification performance for differentiating MCI groups, significantly outperforming neuropsychological metrics. However, instead of purely task-based measures, it was a combination between eye movements and traditional neuropsychological tests that provided the best classification of MCI+

participants overall, although this was not statistically significant compared to other high-performing models.

6.1 Introduction

This thesis has introduced several promising group-level results, especially for biomarker-related eye movements. However, for a new diagnostic test to be considered, it must show comparable or superior classification performance to gold-standard or readily-available tests, especially simple and cheap pen-and-paper ones used in most neuropsychological assessments. A strong example of this has already been demonstrated from a VR spatial task based on hippocampal-entorhinal function: a triangle-completion path integration test provided impressive classification accuracy of MCI+ from MCI- participants when classical cognitive tests performed no better than chance (Howett et al., 2019).

One potential advantage of the iVR task in the current study is the collection of multiple modalities of data simultaneously. Although intimately connected, spatial memory performance and eye movements may together detect more participants with early AD than either one alone. Indeed, there is evidence that a multi-modal approach outperforms a single class of measures in automated diagnostic accuracy for neurocognitive disorders (Knight et al., 2019; Zhang et al., 2019). However, very little research has combined cognitive and eye-tracking modalities for AD diagnosis. Previous studies have demonstrated impressive diagnostic capabilities from visuospatial eye-tracking alone (Parra et al., 2022 ; Crutcher et al., 2009), but these were not directly compared to neuropsychological tests currently used in clinical diagnosis. Demonstrating a *comparative* advantage of spatial eye-tracking measures over current best-practice would provide a strong case to develop them further for clinical assessment.

Hand-in-hand with multi-modal classification of disease is the use of machine learning approaches in medical diagnostics. Availability of computational power and advanced algorithms has provided the means for researchers to build large, complex models for discriminating patients from healthy controls using medical imaging, genomics, natural language and more (see Myszczynska et al., 2020 for applications in neurodegenerative diseases). In particular, super-

vised classification is the branch of machine learning commonly employed in this field, whereby labelled data is used to 'train' predictive models, which are then tested on separated subsets of data to determine classification accuracy. This will be the approach taken in this chapter, which aims to include significant measures from previous sections to answer a critical question for this thesis: can this iVR task provide diagnostic value for detecting early AD.

A key benchmark for this chapter is the comparative classification accuracy of neuropsychological tests currently used in clinical diagnosis of AD. Accordingly, iVR measures were compared against pen-and-paper tests used in Section 4.2.4 for their ability to correctly discriminate between MCI+ and control participants. However, the combination of traditional neuropsychological tests and newer iVR-based measures were also tested for classification accuracy. Indeed, VR measures may outperform neuropsychological measures, but the best classification may involve a combination of the two.

Despite this, a key prior hypothesis for this analysis was that a combined model incorporating both behavioural and eye tracking measures from the iVR task would outperform other models. However, after examining results from the previous two results chapters, we may expect eye movements alone to provide the greatest discrimination, when considering measures from the task only.

6.2 Methods

6.2.1 Participant classification.

Previous studies have examined classification accuracy by comparing MCI+ and MCI- groups only (e.g. Castegnaro et al., 2023; Howett et al., 2019). However, results from the current study have shown less of a difference between healthy controls and MCI+ participants than between the two MCI groups on some variables (e.g. table viewing at encoding [Figure 4.12a] and retrieval [Figure 4.11]; stationary object dwell proportion [Figure 4.10] and seven-second differences [Figure 5.4b]; scan-path vector similarity between encoding and retrieval [Figure 4.13]). Therefore, in addition to biomarker comparison, I included classification analysis of MCI+ participants from MCI- and Older groups combined. This is a conservative approach because healthy older participants were not tested for biomarkers, and were presumed healthy based on ACE-III scores and screening for subjective cognitive decline.

6.2.2 Feature Selection and Engineering

A common issue with statistical learning in medical diagnostics is the high-dimensionality of datasets, where the number of independent variables, or 'features' often vastly surpasses the number of observations. Simply entering all features into a model will invariably lead to poor model performance due to the exponentially increasing feature space, a phenomenon often referred to as the 'curse of dimensionality' (Bellman, 1966). Therefore, the total feature set must be filtered down for high-dimensional datasets. Eye-tracking does not suffer from this problem to the same extent as other areas, but still requires reduction in dimensionality before building classification models.

A common approach to reducing the number of features is principal component analysis (PCA), whereby the dataset is reduced to a new coordinate space with 'components' representing orthogonal linear transformations of the data based on variance explained (Wold et al., 1987). Although this approach

can improve model performance (Odhiambo Omuya et al., 2021; Kheirkhah et al., 2013), it is less easy to interpret because principal components are essentially composed of multiple features. Performing PCA within feature subsets can improve this problem somewhat, especially with high collinearity. Accordingly, principal components of eye movement measurements were included into multimodal classification models, to avoid any detrimental effects of highly collinear eye movement features on model performance. Classification based on these components alone was also shown as a combination eye movement feature set.

Another class of techniques involves selecting the most important features by including a penalisation term for variables that do not significantly improve model performance. There are several types of these 'regularisation' techniques, but most common are L1 and L2 regularisation, which penalise features to zero or linearly, respectively. These can be used to prevent over-fitting of high-dimensional models and provide a means of feature selection (Ng, 2004). For the current chapter, L1 regularisation was used for multivariate models i.e. models with more than one feature. The two most common types of statistical classifier that can include the L1 penalty are (a) the generalised linear model in the form of logistic regression and (b) linear support vector classifiers (SVCs). Both were employed in the pipeline, as described in the next section.

Previous findings throughout this thesis also provide a threshold for feature selection: statistical significance. Analyses from previous chapters have already highlighted measures that show differences between participant groups. Therefore, I focused on features that had either already shown significant differences between MCI+ participants and their control groups, or were relevant to key hypotheses.

In this chapter, the aim was to predict the group label of each participant, rather than quantify the difference between groups. Although previous research has pursued a per-trial classification analysis (Biondi et al., 2018), the goal of a new diagnostic test is to identify individuals with the disease. Accordingly, ag-

gregation of trial-level data to participant-level data is usually required for input into a classification pipeline. The most straightforward approach was to aggregate numerical trial-level metrics by their mean to form an averaged feature per participant. This reduced the amount of data by which to train classification models, but was the most interpretable and relevant approach.

Viewing-level results from Chapter 4.4 were aggregated by their mean per condition, per participant. For time-dynamical data, the overall Ordinary Least Squares fit per participant was used to derive the slope, one second, seven second, and area under the line metrics, ignoring within-participant variation.

6.2.3 Supervised Classification Pipeline

Most classifiers perform poorly when trained on small, imbalanced datasets, such as the the task of identifying MCI+ participants from MCI- and Older participants combined (Althnian et al., 2021). In these cases, models will often default to prediction of the dominant class for each observation, producing misleadingly high accuracy results. A common way to improve model performance in this situation is through data augmentation, whereby the training dataset is artificially increased in size. The Synthetic Minority Oversampling Technique (SMOTE; Chawla et al., 2002) is one such method, which increases the under-represented class by linearly interpolating between observations in feature space. SMOTE was used on MCI+ data when training models, being the minority class in both types of comparisons.

A combination of stratified bootstrap resampling and cross-validation was used to control for over-fitting and estimate statistical significance of model performances. Specifically, the cross-validation procedure was nested within bootstrap resampling, whereby each group was randomly sampled with repetition before testing classification performance. The hyperparameter ‘C’ that governs L1 regularisation was tuned, and the best performing model was tested via stratified 10-fold cross validation for logistic regression and SVC separately. The

maximum performing of these two models was selected as the final result. This pipeline was repeated 1000 times and the resulting distribution used to estimate 95% confidence intervals of model performance.

The classification pipeline was implemented in Python using scikit-learn (Pedregosa et al., 2011) and imblearn (Lemaıtre and Nogueira, 2017) packages.

Evaluating model performance. A popular metric for quantifying classification performance is the area under the curve of the receiver operating characteristic (ROC AUC). The ROC curve is a plot of the false positive rate (1 - specificity) against the true positive rate (sensitivity) for different discrimination thresholds of a binary classification. The ROC AUC provides a single measure of the model's performance across all discrimination thresholds. This metric was taken to evaluate model performance when classifying MCI+ from MCI- participants.

Assessing model performance with an imbalanced dataset was an important consideration for classifying MCI+ from both MCI- and Older adults. ROC AUC will overestimate model performance for imbalanced datasets when attempting to identify the minority class because the number of true negatives will far exceed the number of false positives—by virtue of the imbalanced dataset—leading to an over-inflated specificity.

When predicting a relatively small positive class (i.e. MCI+ participants compared to Older and MCI- combined), a key diagnostic metric to account for is the positive predictive value (PPV; a.k.a precision) of a diagnostic test. This refers to the proportion of correctly predicted positive samples out of all predicted positive samples. A popular way to use this metric is to plot the PPV against the sensitivity of a classification model, with the area under this curve as a measure of model performance. This is similar to the ROC AUC described earlier. However, by effectively swapping specificity for PPV, we can take the false positives into account while avoiding the overinflating effect of the large

number of true negatives from the majority class. The area under the PPV-Sensitivity curve (PS AUC; elsewhere referred to as the 'Precision-Recall' curve, but PPV and sensitivity are used here consistent with diagnostic literature) was used as the main metric for model performance for classifying MCI+ from both Older and MCI- participants.

Note that the scoring is slightly different for ROC AUC and PS AUC. Although both can only have values between 0 and 1, a score of 0.5 in the former indicates a model performing no better than crude or random predictions. By contrast, chance models will tend towards 0 for PS AUC, with scores at 0.5 indicating performance above chance, and scores of 1 indicating perfect separation.

Statistical significance of relative performances from different feature sets was discussed in relation to overlapping 95% confidence intervals.

6.3 Results

Classifying biomarker-positive from biomarker-negative MCI participants. Figure 6.1 shows results of model performances in classifying participants with MCI+ out of all MCI participants (i.e. identifying positive from negative biomarkers) using the area under the receiver operating characteristic (ROC AUC).

The univariate models with the highest ROC AUC used the Stationary Avoidance measure from the Stay-Still condition ($ROC\ \ AUC\ =\ 0.880,\ 95\%\ CI\ [0.809,\ 0.951]$). This measure significantly outperformed task performance ($\%\ Correct\ (Stay\text{-}Still)\text{:}\ ROC\ AUC = 0.656, 95\%\ CI\ [0.560,\ 0.753]$; $\%\ Correct\ (Walk\text{-}Still)\text{:}\ ROC\ AUC\ =\ 0.647,\ 95\%\ CI\ [0.542,\ 0.753]$) and memory monitoring ($Confidence\ Difference\ by\ Correctness)\text{:}\ ROC\ AUC\ =\ 0.683,\ 95\%\ CI\ [0.585,\ 0.781]$).

The stationary avoidance measure also had greater ROC AUC than individual neuropsychological measures, with clearly separated confidence intervals compared to all but the Rey Delayed recall measure, which performed the highest out of all single neuropsychological measures in discriminating biomarker groups ($ROC\ \ AUC\ =\ 0.738,\ 95\%\ CI\ [0.643,\ 0.833]$). However, combining all neuropsychological measures yielded greater mean classification performance than all univariate neuropsychological models, albeit with high variance ($ROC\ AUC = 0.840,\ 95\%\ CI\ [0.720,\ 0.960]$). The combined eye movement feature set using principal components of fixations measures performed at a similar level ($ROC\ AUC = 0.879,\ 95\%\ CI\ [0.786,\ 0.971]$).

Multivariate models with a combination of task modalities significantly outperformed all individual neuropsychological measures ($Task\text{:}\ Memory\ +\ Conf.\ Diff\ +\ Eye\text{:}\ ROC\ \ AUC\ =\ 0.904,\ 95\%\ CI\ [0.844,\ 0.964]$), consistent with hypotheses. However, these did not significantly outperform either all neuropsychological measures together or pure eye movement models.

The feature set with the greatest performance overall included a combi-

nation of neuropsychological and task measures. Specifically, models with principal components of eye measures and neuropsychological metrics had the highest mean and lower confidence limit performance ($PC\ Eye\ +\ Neuropsych:\ ROC\ AUC\ =\ 0.940,\ 95\%\ CI\ [0.885,\ 0.995]$; $Task\ +\ Neuropsych:\ ROC\ AUC\ =\ 0.931,\ 95\%\ CI\ [0.871,\ 0.991]$), although these were not significantly greater than fixation measures or all neuropsychological metrics alone.

A similar relative pattern of results was found for PS AUC scores (Supplementary Figure B.17).

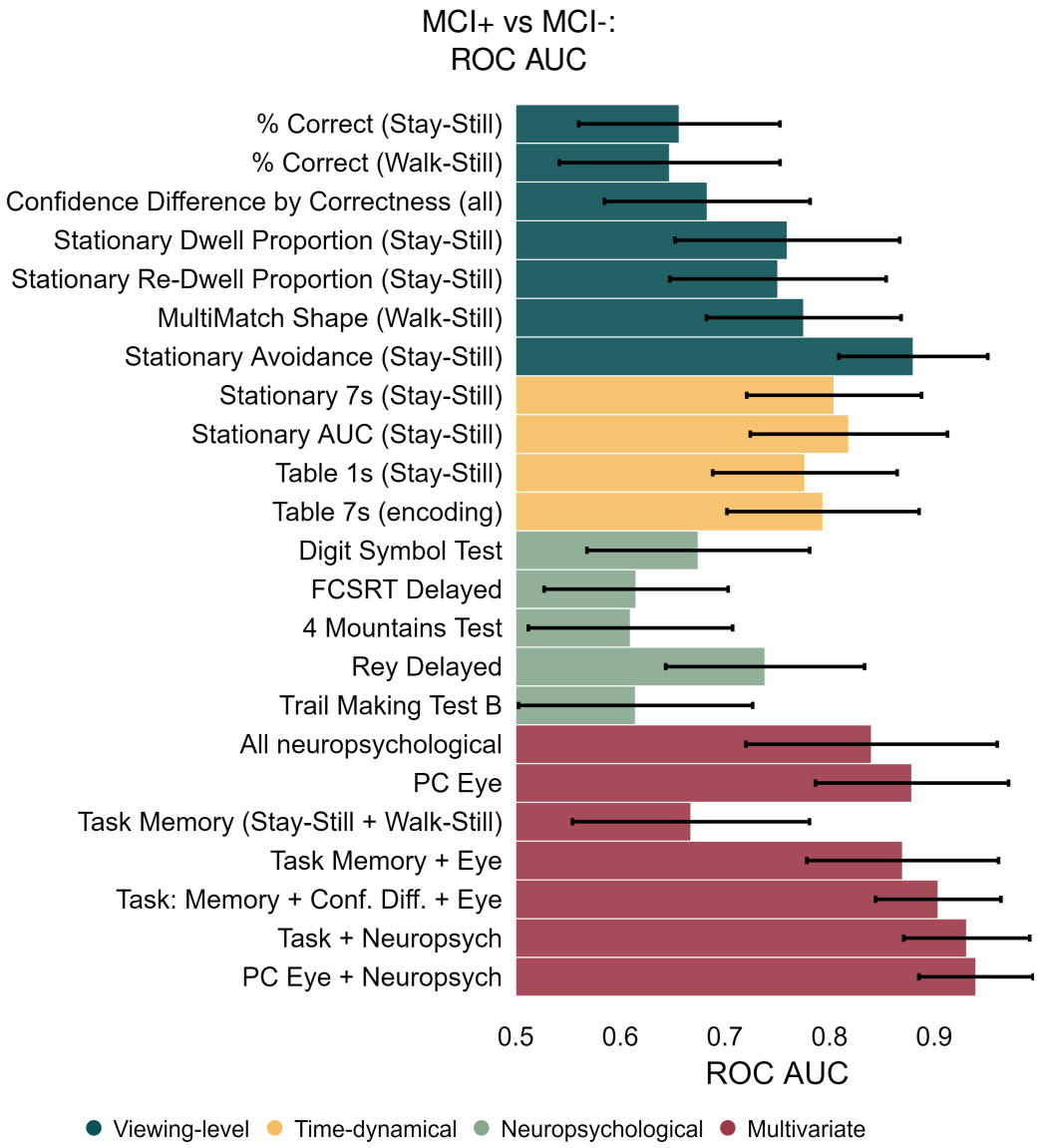


Figure 6.1. Area Under the Receiver Operating Characteristic Curves by feature set for classifying MCI+ from MCI-, coloured by feature category. Error bars are 95% confidence intervals from stratified bootstrapped cross-validation pipeline. Bars are coloured by the type of features the classifiers were trained on.

Classifying biomarker-positive MCI participants from biomarker-negative MCI *and* healthy older adults. Figure 6.2 shows results of model performances in classifying participants with MCI+ out of all MCI *and* Older participants (i.e. identifying positive biomarkers in all age-matched controls). The AUC from the positive predictive value (PPV) plotted against sensitivity (PS AUC) was most appropriate for this comparison (6.2) due to the imbalance of binary class sizes (10:42).

Most confidence intervals were overlapping, indicating that statistical significance could not be reached by the power afforded in this dataset.

The highest mean univariate model performances were Stationary Avoidance in the Stay-Still condition ($PS\ AUC = 0.678,\ 95\%\ CI\ [0.556,\ 0.800]$), the area under the 'curve' (AUC) of stationary object linear time dynamic modelling ($PS\ AUC = 0.694,\ 95\%\ CI\ [0.568,\ 0.820]$), and the Rey Delayed Recall measure ($PS\ AUC = 0.672,\ 95\%\ CI\ [0.550,\ 0.795]$).

For multivariate models, classification of MCI+ participants was similar when combining neuropsychological measures ($All\ Neuropsychological:\ PS\ AUC = 0.770, 95\%\ CI\ [0.631,\ 0.908]$) and combining task measures ($Task:\ Memory\ +\ Conf.\ Diff\ +\ Eye:\ PS\ AUC = 0.722,\ 95\%\ CI\ [0.561,\ 0.883]$). As with the previous section, a combination of principal components of fixation measures and neuropsychological metrics performed the best overall ($PC\ Eye\ +\ Neuropsych:\ PS\ AUC = 0.904,\ 95\%\ CI\ [0.820,\ 0.989]$; $Task\ +\ Neuropsych:\ PS\ AUC = 0.881, 95\%\ CI\ [0.785,\ 0.976]$), significantly outperforming all univariate models. This suggests that, if assessing these participants *prior* to referral for memory assessment, then a combination of neuropsychological and eye movement features may identify participants with positive biomarkers with the greatest balance of PPV and sensitivity.

ROC AUC results from this comparison can also be found in Supplementary Figure B.18.

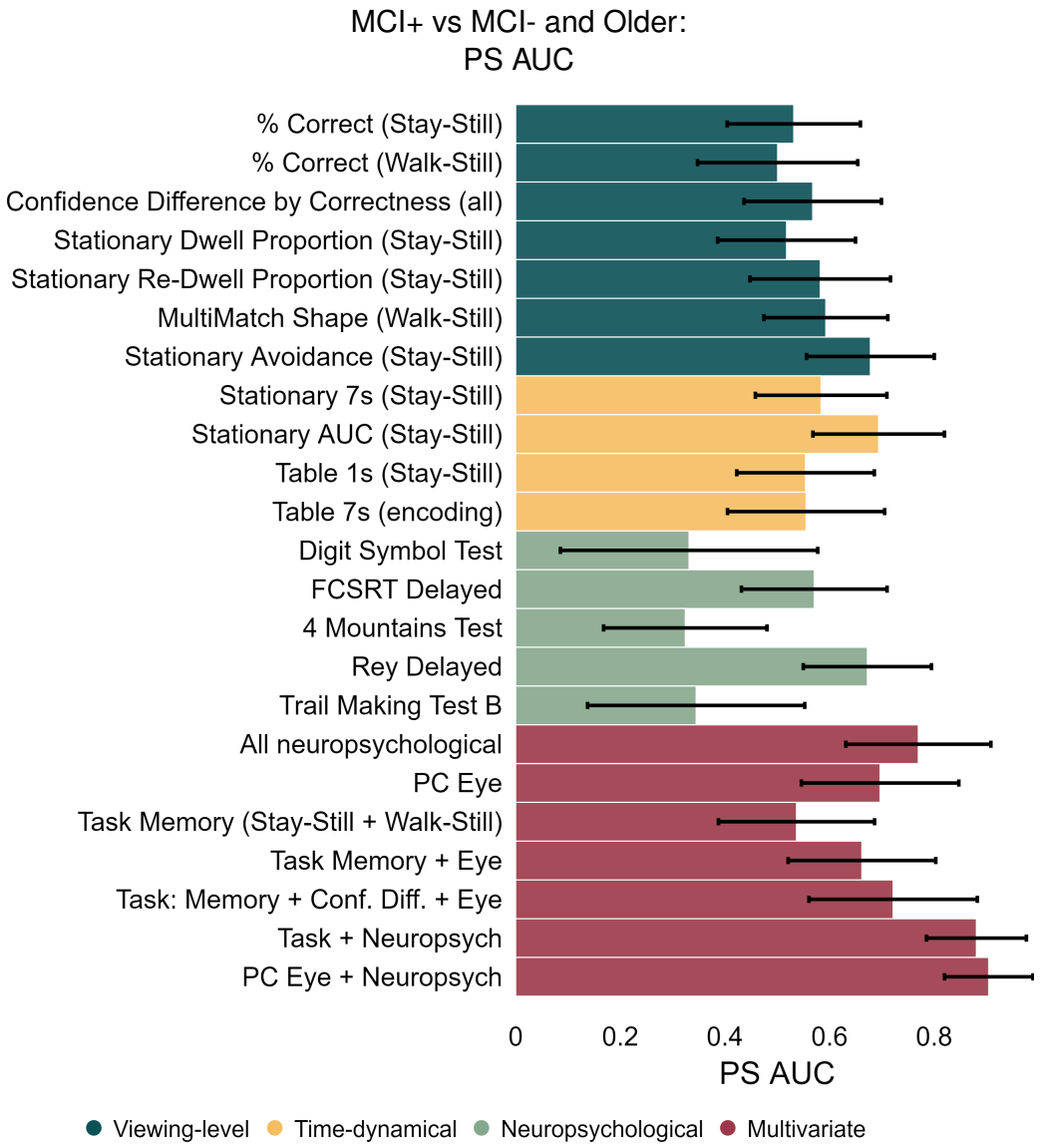


Figure 6.2. Area Under Curves of Positive Predictive Value against Sensitivity aka precision-recall curve for classifying MCI+ from MCI- *and* Older participants. Error bars are 95% confidence intervals from stratified bootstrapped cross-validation pipeline. Bars are coloured by the type of features the classifiers were trained on.

6.4 Discussion

The results presented in this chapter suggest that eye movements during the task may warrant further investigation in aiding detection of early AD beyond the capabilities of more traditional neuropsychological tests. In particular, features derived from the 'stationary avoidance' measure in the Stay-Still condition consistently outperformed pen-and-paper tests, including verbal and visuospatial delayed recall metrics that are (a) commonly used for clinical diagnosis of AD, and (b) have shown a significant difference between MCI groups in this study (shown in Figure 4.3 in Chapter 4). Classifiers that were trained on Stay-Still stationary avoidance had a mean ROC AUC of 0.880 when discriminating MCI+ from MCI-, compared to 0.738 for Rey-Osterrieth Complex Figure (Rey) delayed recall, the highest performing neuropsychological classifier based on a single measure. Although this difference was not statistically significant by the procedure used to estimate confidence intervals, it adds to previous evidence that spatial and eye-tracking measures are effective at identifying AD in it's earliest stages (Castegnaro et al., 2023; Howett et al., 2019; Parra et al., 2022; Lagun et al., 2011), demonstrating potential superiority over pen-and-paper neuropsychological tests in identifying MCI due to Alzheimer's pathology in this sample.

The best discrimination performance overall was achieved when combining eye movement and neuropsychological measures. This was the case for both classification analyses, although these did not significantly outperform other multivariate models, including all neuropsychological measures combined. Still, adding eye movements resulted in tighter confidence limits: there was more confidence that the *lowest* classification performance using both fixation measures and traditional cognitive metrics together was higher than that of purely neuropsychological measures. This suggest that a combined approach could be useful in clinical decision-making for identifying AD, especially from members of the wider population (i.e. identifying MCI+ from MCI- and healthy older).

Multi-modal classifiers based on task memory and eye movement mea-

sures performed similarly compared to combinations between eye movements and traditional neuropsychological tests. Although I hypothesised that a task combination would perform highest overall, finding that a combination between eye movement measures and memory scores performed best at identifying MCI+ participants further supports the notion that these two modalities can be complementary in diagnosis.

Compared to experimental diagnostic tests in other studies, the mean classification performance of the best performing models were comparable to previously published diagnostic tools based on eye-tracking or spatial testing (Lagun et al., 2011; Howett et al., 2019). However, the confidence intervals were wide, likely in part due to the small sample size. Accordingly, results in this chapter should be considered provisional before wider corroboration. Indeed, forward-testing these measures on a future sample would strengthen their relevance, particularly as the stationary avoidance measure was partially devised post hoc.

The supervised learning approach adopted here was relatively simple by modern standards, with room for extension and exploration of techniques, especially those that can train models on more of the available data. For example, previous studies have employed deep learning approaches to use time-series eye data for model training (Sun et al., 2022; Sriram et al., 2023). There are limitations to this type of classification, not least of which is model complexity and difficulty in interpreting feature importance. Another more interpretable approach is the use of Bayesian Networks for diagnostic classification (Seixas et al., 2014). These have become popular in medical diagnosis research due to their inherent ability to incorporate multi-modal data and pre-define qualitative relationships between features.

These are just two examples of methods that could be applied to the current dataset for classifying MCI+ participants. However, the choice of classification technique, while significant, is inherently limited by the quantity and quality of

the data it is trained on. Although eye-tracking allows for a large dataset from relatively few participants, a small sample will still reduce generalisability to new participants (discussion of the limitations of the sample can be found in Section 7.4 of the next chapter).

Nevertheless, I have demonstrated the diagnostic potential of certain components of this iVR task in comparison to current clinical standards. These results will be discussed further alongside other results in the next chapter, with reference to the wider research setting.

Chapter 7

General Discussion

In this thesis, I extended a classic spatial updating paradigm by adapting it to immersive virtual reality (iVR). This adaptation involved incorporating new manipulations of spatial cognition through instantaneous subject transposition ('teleporting'), along with measurement of confidence and eye movements. I aimed to test whether this task was feasible in older adults, and whether spatial memory findings from previous research would be replicated with these participants. I further aimed to test whether this adapted paradigm would provide useful discrimination of participants with early Alzheimer's disease (AD) and therefore demonstrate diagnostic utility. The following subsections describe and discuss the main findings of the thesis, with suggestions for future directions throughout.

7.1 Findings in healthy groups

The adapted task was feasible for testing in older adults. The feasibility of this iVR-adapted paradigm was demonstrated for both younger and older healthy adults based on memory performance, eye-tracking and qualitative results. The number of objects was reduced to improve older adult performance above chance, and the eye data processing pipeline was calibrated and refined. The paradigm could be easily adjusted to change parameters of the task again, such as the angle of viewpoint shift, or the amount of time between viewings.

Although the main objective for adapting this paradigm was to test hypotheses related to AD diagnosis, the task can also be used for future experimentation into spatial manipulations involving the use of physical self-motion.

Indeed, hosting the paradigm in iVR affords several advantages over real-world versions. For example, a current limitation of the task is that the difficulty may be influenced by the specific configuration of objects. However, using iVR to host the task instead of a real-world experiment allows for rapid generation of object configurations with enough variation as to avoid practice effects from similar configurations. Features of object configurations could also be adapted based on current task performance using live feedback into the VR program (as in Heywood-Everett et al., 2022). Additionally, the high level of control of the visual environment removes the need for blindfolds, ear defenders and phosphorescent cues used in previous studies (Burgess et al., 2004), even allowing for impossible manipulations such as teleporting.

Movement by teleportation may be too detrimental to task performance in its current form. Although the inclusion of teleportation was generally successful from a usability perspective, there was some evidence that it made the task more difficult, regardless of table rotation. Specifically, a significant performance- and confidence-reducing effect of teleporting was found for both Younger and Older participants, without an interaction with availability of egocentric visualisation of the array. It is possible that the act of teleporting was disorienting and therefore detrimental to memory performance. Consequently, healthy participants did perform better in the Walk-Still condition compared to Teleport-Still as expected, but they also performed better in Walk-Rotate than Teleport-Rotate. This suggests that the reduction in performance was not due to the disruption to self-motion per se, but due to the process of teleportation itself. This puts into context previous findings of reduced spatial memory performance in teleport conditions compared to walking in healthy older adults (Castegnaro, 2021, ch. 4, p. 149).

Interestingly, this effect was not observed with fixation time on the moved

object, which was otherwise highly correlated with task performance and showed consistent between-condition similarities with percentage of correct trials, at least for healthy participants. Furthermore, there was no evidence that this measure interacted with egocentric or allocentric conditions, and any effect of self-motion was purely driven by the Stay-Still condition. These findings indicate that teleporting had a generally similar effect on moved object dwell as walking did, suggesting that it did not affect this measure by disrupting self-motion.

A limitation of the teleport conditions was that teleporting involved some movement, and could be considered an alternative, more difficult form of self-motion, rather than a lack of it. This is because the act of teleporting in this task required participants to teleport themselves (i.e. they pointed at a target and clicked), and turn towards the table after teleporting. This latter feature allowed for some physical movement to control against the distracting effect of this in walking, but could be considered self-motion towards the new viewpoint.

Future work could remove the post-teleport rotation or develop and contrast different types of VR locomotion to control different spatial processes (Boletsis and Chasanidou, 2022). For example, commercial iVR video games allow customisation of artificial movement between a 'blink' teleport (used in this task) where the screen fades out and back in from a new location, and 'shift' movement involving a fast linear uncontrolled movement to the new location (Valve, 2020; referred to as 'dash' teleporting in Bhandari et al., 2018). Previous research has found that this type of teleportation in iVR reduced spatial disorientation without an increase in motion sickness (Bhandari et al., 2018). Varying the method of instant spatial transposition could allow for research into manipulations of three-dimensional optic flow with and without self-motion. Researchers have already used iVR to demonstrate a reduction in spatial updating memory performance when floor-based optic flow cues were removed (Cardelli et al., 2023). However, iVR has yet to be used for studying optic flow impairments in

AD, despite research suggesting its link to spatial navigation impairment (Kavcic et al., 2006) and potential for diagnostic contribution (Noguchi-Shinohara et al., 2021).

Replication of previous results was mixed and likely confounded. Previous spatial updating results were somewhat replicated by comparing Stay and Walk conditions in healthy younger participants. For example, there was generally an advantage of the object positions being consistent with the egocentric visual snapshot at encoding, as expected. This is the strongest effect seen in previous literature (Simons and Wang, 1998; Burgess et al., 2004). However, Simons and Wang's findings of an interaction between self-motion and visual consistency of the object configuration were not replicated, because seeing the objects with a different retinal projection was not easier with self-motion updating than without (i.e. Walk-Still performance was not greater than Stay-Rotate). The most likely reason for this discrepancy was a key difference in paradigm design: the distracting effect of walking was not controlled for in the Stay conditions in this study, which previously had been achieved by asking participants to walk halfway and then back again during 'Stay' trials. This design decision was made to allow better comparison to the teleport condition, which did not involve walking. However, the same problem applied here in that any disorienting or distracting effect of teleporting was not controlled for when comparing Stay and Teleport conditions. These comparisons were not primary objectives of the study, but closer replication of previous study findings will need to account for this issue.

Older adults had the same patterns of spatial memory results as younger adults but with overall lower performance. The same effects of task condition on memory performance were found in both Younger and Older participants, with Younger participants performing generally higher than their Older

counterparts, as hypothesised. This is consistent with previous findings from other viewpoint-shifting paradigms (Segen et al., 2021; Castegnaro, 2021, ch. 4, p. 149; Hilton et al., 2020), which contrast against navigation studies finding specific allocentric impairments (Colombo et al., 2017; Harris et al., 2012). Indeed, the paradigm may not be sensitive enough to detect spatially-specific ageing impairments discovered by navigation studies.

Alternatively, the deterioration of spatial functioning with age may be explained more by a general age-related cognitive slowing, which affects allocentric conditions in navigational studies due to the increased cognitive effort compared to egocentric strategies (Castillo Escamilla et al., 2023; Markostamou and Coventry, 2022; Wolbers and Hegarty, 2010; Gras et al., 2012). Indeed, Colombo and colleagues suggested that any findings that appear to support specific spatial dysfunctions in older participants may be due to increased load on working memory for the frame of reference that required greater cognitive effort (Colombo et al., 2017). This would explain the results found here and in studies with greater *allocentric* performance in older adults (Castillo Escamilla et al., 2023).

The burden on working memory may have been increased in the current study by the relatively short encoding period, and only one period of learning per trial. Working memory abilities likely had a strong general-purpose effect on performance due to this. A neuropsychological test of working memory was not included, but younger participants did perform significantly faster than older volunteers on Digit Symbol and Trail Making tests, reflecting greater processing speed and executive functioning, which are closely related to working memory (Salthouse, 1992). These results may point to less of a task demand on functions mediated by the medial temporal lobe (MTL) compared to the prefrontal cortex (PFC). Specifically, a general memory impairment in ageing may reflect neurocognitive decline in attentional or working memory functions linked to frontal lobe decline (Pfefferbaum et al., 2005), which affect declarative memory

through the encoding and retrieval of information (Buckner, 2004). This would explain the occurrence of between-condition effects within groups but not between them (i.e. no interaction between age and condition). Research aiming to characterise age-related interactions with spatial memory processes would likely benefit from adapting task difficulty, such as using an adaptive staircase procedure (as in Heywood-Everett et al., 2022).

Eye movement measures discriminated healthy younger from healthy older participants. This was the first spatial updating study to include eye-tracking analysis of viewing behaviour at the second viewing, allowing quantification of spatial memory-related retrieval strategies. Instead of fixations on the moved object predicting age-related memory effects, it was dwell proportions on the table and stationary objects that separated Younger and Older groups.

For table viewing, younger adults had greater table viewing during incorrect trials than correct ones. The reason for this effect is unclear. One possible explanation is that table viewing was a retrieval strategy in younger participants that was halted once the chosen object had been decided. In support of this, linear modelling of time dynamics showed that younger participants decreased their viewing on the table in the Stay-Still condition significantly faster than Older adults after the first second of viewing. Consequently, table viewing in younger adults may have been mediated by uncertainty. Indeed, Younger adults may have utilised more table-based memorisation strategies, as evidenced by increased table viewing at encoding. When they were uncertain of which object moved, they may have tried to identify a memorised shape, or relative positions between pairs of objects. Alternatively, they may have attempted to locate the previous position of the moved object, whereas in correct trials the moved object was identified earlier and table viewing was stopped. These possibilities could be partially investigated in the current dataset by examining the relationship between table viewing time dynamics in high- and low-confident trials.

Regardless, the same effect was not observed for Older participants on average. In fact, healthy older adults appeared to receive a small increase in table viewing with trial correctness. The explanation for this may be consistent with hypotheses: that table viewing was indicative of advantageous memory strategies. Indeed, younger participants—who outperformed older adults—also had greater table viewing in correct trials compared to older adults. Younger participants may have adopted top-down strategies related to table viewing that required more cognitive effort, which older adults therefore avoided (Hess and Ennis, 2012).

An alternative hypothesis is that younger adults were able gather more spatial information using extrafoveal vision (Veneri and Rufa, 2017; Rösler et al., 2005), an ability which is more likely to be reduced with older age (Scialfa et al., 2013; Rösler et al., 2005; Isler et al., 1997). Indeed, peripheral vision has been shown to be important for spatial learning: patients with peripheral visual field loss due to ocular disease had reduced object-location memory performance (Fortenbaugh et al., 2008), and when normally-sighted people had their peripheral vision restricted their spatial recall was more distorted (Fortenbaugh et al., 2007; Alfano and Michel, 1990). Younger participants in the current study may have employed more efficient spatial information gathering by employing extrafoveal vision, helping improve performance.

It may be possible to investigate some of these hypotheses for table viewing in the current dataset by examining scan-path patterns. However, the variation in object configurations likely created too much noise to detect this at retrieval. A more robust test of these effects would keep object configurations consistent across both encoding and retrieval phases for both age groups in each condition.

Table viewing was included as a proxy for a heterogeneous combination of AOIs, which may also explain some of the differences between age groups. For

example, viewing of the previous position of the moved object may be increased in younger adults, which would be consistent with previous results in a change detection paradigm (Yeung et al., 2020). However, the current paradigm did not have salient visual features near the previous position of the moved object, whereas Yeung and colleagues moved an object within a detailed naturalistic scene. Furthermore, identifying fixations that were actually intended for the previous position of the moved object was complicated in the current study by participant and table rotations. Moreover, any overlap between the previous position and (a) the centroid of both encoding and retrieval configurations or (b) the border of the table would need to be accounted for. This may amount to only very few trials with clearly dissociable fixations on the previous position of the moved object from other table AOIs.

In sum, the paradigm was not well-suited to disentangle the different causes of table fixations. However, it may be a useful endeavour to investigate this further to understand age differences in spatial memory processing, contributing to our understanding of age-related cognitive decline.

7.2 AD-related findings

Spatial memory performance did *not* discriminate MCI+ from control participants. Contrary to hypotheses, I did not find a spatial memory impairment in MCI+ participants compared to MCI- participants. No particular impairment was found in the Walk-Still condition, or other conditions with self-motion or allocentric components. One possible explanation for these null findings is that the same cognitive demands of the task that resulted in condition-agnostic memory decline with healthy ageing also masked any specific spatial impairments in MCI participants, which were too subtle for detection with this task. Performance in participants with MCI may have been mostly affected by the demand on working memory or processing speed (Kirova et al., 2015), with any spatially-specific memory requirements being secondary to these cognitive requirements for solv-

ing the task. In support of this, Younger, Older and MCI- participants scored relatively similarly in all conditions except Stay-Still, which involved no distracting movement between encoding and retrieval. It is noteworthy, however, that all participants performed better than chance, consistent with feasibility testing (Chapter 3). Therefore, while there was no floor effect on performance, there may have been a floor effect on the contribution of spatial factors to performance.

An alternative explanation is that spatial memory performance is not as powerful at identifying early AD as previously suggested. In support of this, I found that performance on the Four Mountains Test (4MT) of allocentric spatial memory did *not* discriminate healthy older participants from either MCI group. This fails to replicate previous results of 4MT providing differential discrimination of MCI+ participants from MCI- and healthy controls (Moodley et al., 2015).

Still, there was some evidence of a spatially-specific memory impairment in the MCI+ group when examining performance between conditions. Specifically, MCI+ participants scored as high in the Walk-Rotate condition as the Stay-Still condition. Indeed, if we assume that teleport conditions confounded the difficulty of the task, as discussed earlier, then task performance was almost exclusively moderated by the availability of an egocentric visual 'snapshot' of the object positions for this group. This supports the hypothesis of an impairment in self-motion and allocentric processing for spatial memory in early AD.

Although this finding is consistent with hypotheses within MCI+ participants, it is unexpected in relation to results from MCI- participants. Of course, in such a small sample, one possibility is that this was a spurious finding. However, it could also be consistent with the notion that performance can be explained by demand on both working memory and spatial memory. The MCI- group may have been particularly susceptible to the former because their cognitive impairment was caused by an age-related decline in this area of functioning, rather than Alzheimer's pathology. Conversely, the MCI+ group may—probably

additionally—have had specific allocentric and self-motion impairments, leading to an over-reliance on egocentric strategies by default. Conversely, MCI- volunteers could have had conflict between spatial processes that reduced performance in Move-Rotate conditions.

This pattern of effects may have contributed towards the seemingly *higher* performance in the MCI+ group in the Walk-Rotate condition compared to MCI- participants: it is plausible that some MCI+ participants even had less impaired working memory than the MCI- group, receiving a performance boost in egocentric conditions in which they were less spatially impaired. Indeed, some studies have found intact working memory performance in MCI participants that later converted to AD (Lee et al., 2014). Of course, this does not explain the relatively low performance of the MCI+ participants in the Stay-Still condition, and also is not consistent with literature suggesting *additional* working memory deficits in MCI caused by AD compared to other sub-types (Belleville et al., 2007; Saunders and Summers, 2011).

Regardless, no conditions discriminated biomarker groups effectively by pairwise comparison or classification analysis. This is in contrast to studies finding reduced performance specific to spatial conditions in MCI participants at elevated risk of AD. For example, one study found that participants with amnestic MCI were impaired on spatial reference memory from a virtual radial arm maze, predicting later conversion to AD (Lee et al., 2014). These same participants were not impaired in working memory tests compared to age-matched controls. This paradigm may have been more sensitive to AD-related MTL damage than the current spatial updating task due to the navigational element of the radial arm maze. Similarly, the path integration test (PIT), developed in the same research group as the current work, was designed to place demand on entorhinal cortex navigational functions rather than object-location memory, finding a strong discriminating effect between MCI+ and MCI- participants (Howett et al., 2019). Another study from our group focusing on object-location memory did

not find the same effect (Castegnaro et al., 2022) suggesting that tasks involving memory for object locations may be less sensitive to AD neuropathology compared to spatial navigation paradigms.

Research on neural correlates of spatial memory and navigation implicate entorhinal and hippocampal networks in both functions (Burgess et al., 2002; Hartley et al., 2014; Bush et al., 2014; O'Keefe and Burgess, 2005), which are tightly inter-related. However, research suggests different processing streams for object-identity and content information in the anterior-temporal MTL compared to posterior-medial areas for spatial-context coding (Reagh et al., 2018; Maass et al., 2015). Although there is evidence for impairment in both areas in AD patients (Khan et al., 2014; Braak and Braak, 1991), the anterior-temporal regions may also become dysfunctional with healthy ageing. Indeed, Reagh and colleagues found antero-lateral EC activity was related to impaired object pattern separation in healthy older adults, but not on a spatial task (Reagh et al., 2018). Therefore, tasks that rely on object identity processing may not discriminate early AD from age-matched controls because there is also an age-related confound.

Alternatively, object-location tasks, such as this one, may allow for non-spatial strategies that are not available in self-location tasks. For example, participants can use non-spatial features of objects combined with flattened spatial heuristics (e.g. remembering colours in a clockwise order) as a mnemonic strategy in the current task, which may rely more on working memory functions. One option to counteract this effect could be to use identical objects (as in Segen et al., 2020). However, this would remove the possibility of swap errors, which have been greater in familial AD participants in a visuospatial object-location binding task (Liang et al., 2016). The use of semantically-neutral visual stimuli in Liang and colleagues' paradigm (described in more detail here Pertzov et al., 2013) may also avoid the use of compensatory strategies that involve categorising or taxonomising objects (Postma and De Haan, 1996).

A potential avenue for improving the diagnostics of spatial memory performance in this task could be to focus on recall of object locations after a delay of several minutes, rather than a few seconds. Reduced memory performance with such a delay is not just a well-established impairment in AD (Estévez-González et al., 2003; Guarch et al., 2008), but the most common method of clinical diagnosis (Cerami et al., 2017). This is illustrated in the current thesis by the delayed recall metric in the Rey-Osterrieth Complex Figure Test, showing the greatest discrimination of MCI+ from control participants of all comparator neuropsychological tests. Delayed recall tests were chosen in this study for their clinical use in diagnosis of AD, based on evidence of predictive validity for conversion (e.g. Derby et al., 2013; Corwin and Bylsma, 1993). However, spatial memory has rarely been examined with delays of 25 minutes or longer like in clinical delayed recall tests, with existing results purporting to study spatial cognition relying more on visuo-spatial abilities (de Toledo-Morrell et al., 2000; Flicker et al., 1993; de Toledo-Morrell et al., 1984). The combination of allocentric spatial memory and recall delay may increase the demand on impaired cognitive functions in AD patients than either one alone. A potential option for this could involve a complex one-trial approach where participants would learn numerous object locations and replace them later (similar to Postma and De Haan, 1996).

Eye movement measures *did* provide useful discrimination of MCI+ participants from other groups. The best way to discriminate MCI+ from control participants in this study was with eye fixation metrics. Several measures discriminated MCI+ from MCI- and Older participants with superior accuracy than comparative neuropsychological tests, including delayed recall scores. These measures were mainly derived from viewing behaviour on the stationary objects, consistent with but not exactly the same as hypotheses.

The hypothesised measures based on fixations on the moved object were not themselves effective discriminators. However, the *change* in likelihood of

viewing the moved object over time was significantly lower for the MCI+ group compared to all others, even if total viewing time on the moved object was not. This latter finding is in contrast to results on the visual paired comparison (VPC) task, which showed that percentage of time spent looking at novel stimuli was predictive of conversion from healthy control to MCI to AD (Zola et al., 2013). The version of the VPC task used by Zola and colleagues did not involve declarative memory: participants were just required to look at stimuli 'as if watching television'. This is consistent with earlier results showing that participants with AD viewed novel and irregular visual stimuli less than controls during an exploratory eye-tracking paradigm (Chau et al., 2017, Daffner et al., 1992).

Although the current task involved a novel change to the visual stimuli, the paradigm was more task-oriented, which may have resulted in different viewing behaviour than an exploratory task. Alternatively, the short viewing period may have reduced the power of a percentage viewing measure, which a time-dynamical measure could detect instead. Indeed, if the linear models of moved object gaze time dynamics were projected further in time for the Stay-Still condition, then the areas under the lines would presumably become statistically different. However, if the slope of this trend is more discriminatory than raw dwell percentage, other change detection or novelty paradigms (including VPC) that have found effects with viewing-level measures may find an even larger effect by analysing the data in a similar way.

No differences were found in viewing of the moved object in allocentric viewpoint-shifting conditions between MCI+ and MCI- participants (Walk-Still and Teleport-Still), consistent with task performance but contrary to hypotheses. This was the first time eye movements were examined in this spatial updating task at retrieval. However, previous research has found that a moved object was not fixated on more by healthy younger adults after a viewpoint shift (a movie scene 'cutting' to a new viewpoint with a 90° shift) than if the object did not move (Hirose et al., 2010). This suggests that these allocentric viewpoint-

shifting tasks do not have the same preferential viewing effect on eye movements as other change-detection tasks. Indeed, when the viewpoint is rotated, all visual information has shifted, so a detection of a change must be made relative to a change of the whole scene. This may disrupt the usual preferential viewing effect, as visual cues must be related back to allocentric or relative object locations.

One eye movement measure did show a significant difference between MCI+ and control groups in the Walk-Still condition. The MultiMatch 'shape' metric, which compared sequences of saccade vectors between encoding and retrieval viewing periods, was particularly low for MCI+ participants in the Walk-Still condition, suggesting a specific effect of this condition on the familiarity of scan-path patterns between encoding and retrieval. This may be a reduced gaze reinstatement effect, whereby saccadic vectors at encoding are reproduced at retrieval less than other conditions or groups.

To my knowledge, no research has directly investigated gaze reinstatement in AD, let alone under a viewpoint change. Neural correlates of gaze reinstatement have been mixed, with recent evidence showing no significant prediction of reinstatement from activity in the MTL (Wynn et al., 2022), despite models of gaze reinstatement implicating areas of the MTL involved in spatial cognition (Bicanski and Burgess, 2019). This thesis might provide some early empirical support for this theory, assuming that it is MTL-related damage in MCI+ participants driving the effect.

Interestingly, the reduced vector sequence similarity between encoding and retrieval for MCI+ participants was potentially specific to *correct* Walk-Still trials. This is opposite to the hypothesised effect, whereby signs of gaze reinstatement were predicted to be associated with greater performance (Wynn et al., 2022). However, if this effect is real, then it cannot be simply related to performance because (a) MCI+ participants had an overall positive correlation between performance and vector sequence similarity, (b) we do not observe the same effect

in other conditions with similarly poor performance, such as Teleport-Still, and (c) vector sequence similarity was not associated with performance in Younger, Older or MCI- groups. Further interrogation of this effect is required to test its robustness, but MCI+ participants may be employing less gaze reinstatement during Walk-Still trials specifically, which is exacerbated in correct trials due a different retrieval strategy.

Despite findings of group differences in the Walk-Still condition, it was the Stay-Still condition that showed the most consistent eye movement changes in the MCI+ group compared to control participants. For example, the fixation time on the stationary objects was greater at retrieval in Stay-Still specifically. Subtracting this from dwell proportions on the table and the moved object gave relatively strong discrimination of MCI+ participants. This supports aforementioned findings of reduced preferential viewing effects in AD for static scenes without viewpoint shifts (Ryan et al., 2020).

One reason for this finding may be an effect of the aforementioned task difficulty. MCI+ participants performed somewhat lower than other groups in the Stay-Still condition, although not significantly when compared to MCI- participants. Eye movement metrics may simply have been a more powerful metric to detect this difference, and the Stay-Still condition provided the best-calibrated memory demands for observing its effects on eye movements. Indeed, task performance in other groups suggests that this condition was much easier than others. However, the most discriminating eye metrics were based mainly on table and stationary viewing in the Stay-Still condition. These were not as closely associated with task performance as fixations on the moved object, although all main fixation measures showed differences between correct and incorrect trials in the MCI+ group like other groups. Still, eye fixation measures discriminated MCI+ participants much better than task performance in classification models, supporting their use in detection of AD.

Participants in the MCI+ group showed a distinct time-dynamic profile of fixation probabilities during Stay-Still retrieval phases. The lesser increase in moved object viewing compared to other groups has already been mentioned in this section. In addition, MCI+ participants were less likely to view the table *early* in Stay-Still retrieval phases, whereas all groups tended to converge by the end of the seven seconds. Although this was not the most discriminating measure—metrics derived from stationary object viewing performed best here—this early avoidance of table viewing was distinct in being an effect that did not enlarge over the course of the viewing period. This suggests that MCI+ participants may be reacting immediately differently to retrieval stimuli than other participants. Linear time-dynamic analyses removed the first second of eye tracking data, but this early data could hold useful insights into stimuli reactions, and could be modelled using a non-linear approach (such as seen in Oleson et al., 2017).

Results suggest that MCI+ participants may have had a similar pattern of eye behaviour across conditions, whereas other participants changed their viewing behaviour in the Stay-Still condition compared to other conditions. Indeed, the MCI+ participants were the only group to not have a different proportion of dwell on the moved and stationary objects in the Stay-Still condition compared to other conditions. Furthermore, MCI+ participants did not show changes to dwell proportions from encoding to retrieval in the Stay-Still condition like other participants did. These findings are consistent with previous results showing that AD participants did not adapt their viewing to task requirements as much as healthy controls (Shakespeare, Pertzov, et al., 2015). Specifically, Shakespeare and colleagues showed that healthy controls shifted their visuospatial attention when asked to search for visual features of scenes, compared to scanning the scenes without a search task. However, patients with AD showed very little difference between task and non-task scan-paths. Although the search task was not a memory task per se, the authors proposed that eye movement results

found in AD patients were due to memory impairment. The same type of effect may be present in the current study.

For encoding-specific eye movements, group differences were moderate, with the strongest effect showing less discrimination of MCI+ participants compared to eye movements at retrieval (i.e. Table 7s at encoding in Figures 6.1-6.2). This is somewhat unexpected because previous studies have found changes in stimuli fixation times during encoding phases in AD patients (Pavisic et al., 2021; Parra et al., 2022). Furthermore, no encoding entropy differences were found, contrary to results from previous MCI (Coco et al., 2021) and MTL damage (Lucas et al., 2018) studies. One explanation for these findings in the current task may be that encoding patterns were heavily influenced by the configuration of objects. Specific configurations may have led to different encoding strategies, measurable through fixation patterns. A further exploration of this issue can be found in Appendix A.

Confidence rating may discriminate MCI participants from healthy volunteers, but not be sensitive to presence of AD biomarkers. Although all groups had higher confidence rating in correct trials compared to incorrect ones, MCI participants had less of a difference here than healthy groups and also did not have a significant association between overall averaged confidence and memory performance, unlike Younger and Older participants. This suggests an impairment in spatial memory monitoring may exist for those diagnosed with MCI. However, a specific relationship to AD could not be detected here.

All MCI participants, including those with negative biomarker results, would have approached clinical services due to subjective memory problems and been subsequently assessed using the Addenbrooke's Cognitive Examination (ACE-III; Mathuranath et al., 2000). MCI- participants may, therefore, represent a group of healthily ageing adults with more detectable age-related memory decline than participants in the healthy Older group. It is possible that reduced

memory monitoring abilities is correlated with age-related cognitive decline or cognitive abilities in general. This is supported by evidence for more high confidence errors in older adults compared to younger participants (Dodson et al., 2007; Shing et al., 2009; Angel et al., 2022). Interestingly, research suggests that the vividness of subjective memory experiences are the same or even increased with age, despite reductions in objective accuracy (Johnson et al., 2015; Comblain et al., 2004). These findings suggest that younger and older adults may weight mnemonic features differently when encoding or retrieving memories (Johnson et al., 2015). Future research could test whether this also applies to spatial memory paradigms by varying spatially-contingent emotional vividness of stimuli.

On the other hand, spatial memory monitoring may not be the best candidate for inclusion in diagnostic testing of AD based on the results in this study. This is consistent with previous mixed findings in metamemory research, with one study showing that AD patients had intact local, task-based memory monitoring despite global anosagnosia (Gallo et al., 2012) and another showing preserved recognition memory monitoring (Moulin et al., 2003). Still, it would be a potentially informative area of research to combine two confidence scales for two different sub-tasks of the spatial updating paradigm: one after object selection, and another after a new object replacement phase (discussed below). Memory monitoring for spatial coordinate estimation may be different from spatial recognition in such a change-detection paradigm. This would also allow for a comparison of the difference between two confidence scores, which may normalise some individual differences in self-reporting tendencies.

7.3 Implications for diagnosis of AD

The impact of these results is necessarily limited by the participant sample collected to date. Although eye-tracking appears to provide the same power with fewer trials than some behavioural measures, the ability for a diagnostic test to

generalise relies on a representative sample, otherwise there is a risk of overfitting a test to noise, particular with data-rich techniques. Not only were MCI participants few in this study, they were also non-representative in ethnicity, sex, and culture. Efforts have been made to increase awareness of the lack of diversity in clinical research samples (Bhopal, 2008; Bartlett et al., 2005), which can lead to medical exclusion or increased risk of serious adverse events (Bartlett et al., 2005). Accordingly, the results in this thesis are provisional, received from a convenience sample during an extremely reduced data collection time-frame. However, receiving participation from any MCI patients is valuable, especially those with biomarker testing. Therefore, despite these caveats, the potential implications of this and similar research on the field of AD diagnosis will be discussed, with the significant proviso that it is unknown whether the results of the study would generalise to a larger, more heterogeneous group.

Evidence for the use of an eye movement-based diagnostic test of AD is building support (Readman et al., 2021; Wilcockson et al., 2019; Beltrán et al., 2018; Fernández et al., 2015), with this study making the significant contribution of directly comparing MCI participants with and without biomarkers for AD and finding differences in eye movements with high diagnostic potential. Moreover, I have demonstrated that eye movement measures were more accurate for diagnosing this sample of MCI+ participants than verbal, visuospatial or spatial memory tests. These findings lead to two important questions: (1) is this likely to be the best eye-tracking test of AD, and (2) are eye-tracking diagnostic tests better candidates for clinical translation than alternative approaches?

Although the eye movement findings in this study are promising, other eye tracking tests have previously made highly accurate predictive classification of AD patients. For example, one study achieved 100% specificity and 94% sensitivity when predicting which of 65 MCI participants would convert to AD using eye movement and memory measures of a visual short-term memory binding task (Parra et al., 2022). This required a simple computer monitor and 10 min-

utes to complete. By contrast, the current task lasts about 15 minutes for only two conditions, requires more expensive equipment in a large space, and so far has demonstrated inferior diagnostic metrics from a smaller sample. The main advantage is its integration with an immersive head-mounted display, which is advantageous insofar as the technology can demonstrate superior diagnosis to more convenient alternatives. Yet, the spatial memory testing from the new task is not as diagnostically accurate as eye movement measures, or previous spatial tests such as the PIT (Howett et al., 2019). Furthermore, much simpler eye-tracking paradigms based on simple fixation, saccade or anti-saccade tasks have been shown to discriminate between AD and other neurodegenerative conditions (Opwonya et al., 2022; Shakespeare, Kaski, et al., 2015). Therefore, the use of this task for further diagnostic testing in its current form is, in my opinion, only justified if it can be combined with a behavioural or other marker that affords a significant advantage over alternatives.

One potential direction for this might be adapting the spatial updating task to include object replacement, creating a metric of spatial error per trial that may be more sensitive to AD-related changes than the current object selection. Indeed, there is some evidence that testing spatial coordinates for object location memory is more sensitive to AD than coarser methods (Ruggiero et al., 2020). However, this potential adaptation to the paradigm is not without its complications: for example, some approaches allow visualisation of object replacement (e.g. Cherry and Park, 1993; Pertzov et al., 2013; Castegnaro, 2021, ch. 4, p. 149), potentially introducing a visual recognition component to the task. An alternative method could involve visual feedback only after positional judgement (as in Doeller et al., 2008), although if replacing objects by virtual pointer then some error would be due to motor control. Nevertheless, object replacement could be an interesting option for future research as it may also affect participants' eye movements if the task focuses on 'where' as well as 'which'. One hypothesis is that this would result in more fixation time on the table as participants decide the

moved object's previous position. Results from this thesis suggests that MCI+ participants may have reduced table viewing relative to controls with this task adaptation.

Alternatively, more success might be achieved by combining eye tracking with behavioural or cognitive diagnostic tests that have already demonstrated high accuracy for identifying AD. For example, spatial tasks such as PIT could examine gaze-based navigation strategies using VR-eye-tracking. Indeed, experimental paradigms have already included gaze tracking as a tool to study navigation abilities (e.g. Bécu et al., 2019; Enders et al., 2021; Drewes et al., 2021). Alternatively, standardised neuropsychological tests could include eye movement analysis with desktop eye-tracking, which has already been achieved for visuospatial figure-copying (Kim et al., 2022), trail-making (Hicks et al., 2013) and pattern completion (Hayes et al., 2011). However, a danger of adding eye-tracking to memory tasks, including the one in this thesis, is the risk of over-fitting to noise, especially in small samples. Care must be taken to set clear, justified hypotheses and forward-test any post-hoc findings on new samples.

Of course, non-eye-tracking candidates exist for improving early diagnosis of AD. The aforementioned PIT spatial navigation task yielded 92% sensitivity and specificity for MCI+ participants compared to MCI- controls (Howett et al., 2019), building upon a strong precedent set by prior findings of AD-related spatial navigation impairments (Hort et al., 2007; Weniger et al., 2011; Serino et al., 2015; Coughlan et al., 2018). Beyond spatial testing, spontaneous speech processing has yielded impressive classification accuracy for AD (Qi et al., 2023). Furthermore, improvements in episodic memory testing of AD have also been found more recently by incorporating a memory binding component (Pereira et al., 2023).

Additionally, neuroimaging studies have seen an increase in the application of deep neural networks to magnetic resonance images (MRI) for detection of early AD brains (Kehoe et al., 2014; Yamanakkanavar et al., 2020; Mehmood

et al., 2021). Moreover, the identification of AD biomarkers through blood testing has shown highly promising results (Varesi et al., 2022; Gao et al., 2023) with methods such as oral samples (e.g. saliva) providing non-invasive alternatives (Paraskevaidi et al., 2020).

The benefits of cognitive and behavioural tests over pure biomarker definitions of AD were discussed in the introductory chapter. Briefly, relying on protein thresholds alone produces false positives from individuals with existence of amyloid and tau but no cognitive or behavioural effects. Cognitive markers also provide clinically-meaningful outcome measures for intervention effectiveness in clinical trials, and more generally provide a neuropsychological metric that may closely relate to clinical symptomology. Additionally, VR and eye-tracking tests would have the advantage of being relatively less costly than neuroimaging, less invasive than CSF or blood testing, but more accurate to early changes than pen-and-paper tests. However, their ability to scale to national or international healthcare systems is as-yet unproven, with a number of significant barriers preventing smooth adoption, such as physical space requirements, technical support, and cultural adoption. Especially in non-specialist clinical settings, the practical advantage of a quick pen-and-paper test far outweighs the accuracy benefits from a diagnostic test requiring expensive, bulky and error-prone technical equipment. This point applies to eye-tracking in any form, particularly if their diagnostic accuracy relies upon a prohibitively large number of trials.

Still, building an evidence base for immersive diagnostics will allow swift implementation of effective tools should the use of mixed reality technologies become widespread throughout healthcare and wider society. They may never replace quicker and more convenient techniques, but could provide cost, time and acceptability advantages over specialist tests such as scanning.

7.4 Limitations and considerations

The spatial updating paradigm. Some key elements of the spatial updating paradigm were changed from previous versions for the research in this thesis. For instance, the trial conditions were not communicated to the participants beforehand. Instead, the instructions to walk, stay or teleport were made after the first viewing, and the rotation of the table (or lack thereof) was made visually obvious before the second viewing. These changes were implemented following early piloting when it became clear that even healthy younger participants could not remember the condition of the current trial. We agreed that it would not affect the reliance on allocentric memory or self-motion in key conditions, and therefore would still be a suitable task for the hypotheses and potential diagnostic targets. It also allowed for potential scan-path clustering at encoding (see Appendix A), because each encoding period was condition agnostic. However, it is possible that introducing this change may have affected the performance of the task, the most intuitive effect being a reduction in memory performance, assuming that there were more effective strategies for some conditions than others. Additionally, it is feasible that encoding eye movements would be different if the participant knew they would be retrieving the information from the same or different viewpoint as the first viewing. The effects of this could be studied in future work, although the difficulty of the task should probably be reduced as extra instructions would add to cognitive load, especially for older adults.

Another key change was the use of exclusively 135° table rotations and viewpoint shifts to minimise the availability of egocentric strategies in shifted conditions. However, this also increased the difficulty of the task (Mou and McNamara, 2002; Heywood-Everett et al., 2022), which may have masked spatial memory impairment in MCI+ participants. Although feasibility testing identified a need to reduce the number of objects from five to four, future research could further reduce this to three objects to maintain spatial reference frames (i.e. low availability of egocentric strategies with 135° viewpoint shifts) for key conditions

while increasing performance in MCI- participants. In support of this, previous research has used a similar paradigm with only three objects, finding AD-related differences in spatial memory (Ruggiero et al., 2020). Indeed, reducing the difficulty of Walk-Still trials to match healthy performance in the Stay-Still condition may allow closer testing impairments specific to MCI+ participants.

A lesser-addressed consideration in spatial updating paradigms is the interaction between spatial frames of reference and the configuration of objects themselves. Iachini, Ruggiero and colleagues developed a paradigm that came closest, by controlling the relative egocentric and allocentric inter-object distances, then probing egocentric and allocentric understanding based on categorical or coordinate judgements of object locations (Ruggiero et al., 2020; Ruggiero et al., 2018; Ruggiero et al., 2012; Iachini and Ruggiero, 2006). However, a similar control of inter-object distances from different frames of reference has not been studied with viewpoint shifts or object movement. One potential direction could adapt the current paradigm to vary the egocentric consistency of the *moved* object despite a viewpoint shift, by translating its world-space location in such a way as to be in the same egocentric position as encoding, even after a viewpoint shift. This could test egocentric-allocentric switching, which Ruggiero and colleagues' research suggests may be impaired in AD (Ruggiero et al., 2020; Ruggiero et al., 2018).

Eye movement considerations. Although lowering the difficulty of the task has been discussed, an adaptive-difficulty task may be useful for specifically exploring eye movements in healthy and patient groups. This is because eye movements appear to be mediated by trial correctness, but the paradigm currently does not counterbalance task difficulty, resulting in a varying number of correct trials by condition per participant, and by group. Task difficulty could be calibrated per participant to improve the balance of correct and incorrect trials. For example, the amount of encoding or retrieval time could be used as the

parameter for task difficulty, with average encoding time saved and compared across groups. Other parameters for task difficulty could be used, such as number of objects, spatial configuration of objects, or object shift distance (Heywood-Everett et al., 2022). Regardless, differences in correctness-dependent fixation patterns could then be compared across groups more robustly, potentially providing a powerful means for detecting gaze differences in AD.

A key limitation of the eye movement analysis in this thesis was the avoidance of saccadic measures. The precision and frame-rate of the eye-tracker were decidedly too low to measure onset, offset and velocity of saccades that are required for certain measures such as saccadic over- or under-shooting. This is despite pro-saccade and anti-saccade studies often finding abnormalities in eye movements of patients with AD and MCI (see Opwonya et al., 2022 for a review). Furthermore, the study of microsaccades requires even greater resolution, but has successfully detected differences in AD and MCI patients (Kapoula et al., 2014). However, saccade effects are less often studied in free viewing and naturalistic tasks such as this one due to their unconstrained nature. The addition of free head movements in VR settings only adds to this issue. Still, incorporating sensitivity to saccadic differences in a spatial memory task may add to its diagnostic value.

An additional consideration is the inability of the current paradigm to detect oculomotor dysfunction. The eye-tracking hardware and data processing pipeline were not configured to detect such phenomena as square-wave jerks, which have been found to be more prevalent in AD participants than age-matched controls in some results (Shakespeare, Kaski, et al., 2015; Kapoula et al., 2014), although not in others (Pavisic et al., 2017; Zaccara et al., 1992). The ability to distinguish between these and technical eye-tracking drop-out was not present with the materials and setup of the study. It is possible that the presence of such oculumotor dysfunctions reduced the accuracy of foveated gaze in the

MCI+ group, a notion that is supported by AD-related impairments in saccade and anti-saccade tasks (Opwonya et al., 2022). Such inaccuracies could have biased fixation results, such as the amount of dwell time on the table. Future studies of this paradigm should include tests of oculumotor function to control or correct for any potential confounding here.

7.5 Methodological contributions

Aside from the iVR task itself, this thesis has introduced a number of novel methodological contributions to the field. Firstly, I have demonstrated the benefit of observing eye movements during retrieval in the spatial updating paradigm. By including a period of forced viewing prior to object selection, fixation patterns and statistics can be compared and contrasted between conditions and participants. Admittedly, there may be a way of incorporating eye tracking at retrieval without this forced viewing length, which required participants to wait even if they had mentally selected the chosen object before the end of the viewing period. Previous paradigms avoided retrieval gaze analysis because viewings with fast reaction times were too short for this (Segen et al., 2021, Hilton et al., 2020). One potential solution to this is providing a task which cannot be solved quickly. Although I have suggested reducing the number of objects further to make the task easier for MCI- participants, an advantage of increasing the number of objects is that it would likely require longer viewing times at retrieval if a forced viewing period was removed. This approach could allow for modelling of decision-making and choice processing (Thomas et al., 2019; Tavares et al., 2017), which may add to detection of early AD (Laurens et al., 2019).

Removing the fixed viewing period would likely invalidate the approach to linear time dynamics demonstrated in Chapter 5. Indeed, this way of modelling fixation-time data is probably often avoided because it is not the best fit to the paradigm. However, the linear trend in probability data over time is a useful finding in itself in this study because it allows for an easy-to-interpret and rela-

tively simple way of examining changes in viewing behaviour at the group and participant level.

The analyses of gaze changes over time in Chapter 5 was just one of several potential analyses that could have been included in this thesis. Another method could examine cumulative number of fixations at each time point for single or multiple AOIs (e.g. moved object or all configuration objects) and fit non-linear curves. Additionally, fitting non-linear curves to the first second of retrieval viewing may reveal differences in reactions to the changed stimuli, as mentioned earlier. Alternatively, an approach to clustering scan-paths at encoding was developed for this thesis, but was not appropriate for the current dataset due to randomisation procedures in the study (Appendix A). Still it exemplifies the breadth of methodologies for analysing eye movement data, and can be used on future datasets to investigate mnemonic strategies through gaze patterns, extending previous spatial updating findings (Hilton et al., 2020).

Many other time-dynamic approaches could be devised, which speaks to the power of eye-tracking and the amount of information it provides. It is noteworthy that several differences in MCI+ gaze patterns were discovered from time-dynamic analyses, supporting the use of time-dependent measurements in eye movement analysis, which are often overlooked for simple cross-sectional measures.

7.6 Conclusions and final considerations

This thesis has shown that a spatial memory task with eye-tracking can provide high classification accuracy of MCI patients with positive biomarkers from control participants based on eye movement measurements. This includes superior performance compared to neuropsychological tests used in clinical practice, similar to those considered the gold-standard for memory assessment. This adds to the evidence that these conventional tests could be augmented or even superseded with more accurate tools, although future diagnostic tools

must prove their ability to fit into established healthcare systems.

A key finding from the thesis was that the allocentric walk condition was not as discriminating as the relatively more straightforward egocentric condition. Admittedly, if a spatial walk condition is not as diagnostically accurate as a simple egocentric task, then iVR may not be needed. Accordingly, one could develop a desktop, laptop or even smartphone version, allowing greater reach and reducing expense. Indeed, previous spatial tasks with sensitivity to AD have been hosted on 2D screens (Serino et al., 2015; Weniger et al., 2011). Furthermore, smartphone-based eye-tracking also now exists, showing impressive validity when compared to head-fixed setups (Valliappan et al., 2020; Gunawardena et al., 2022). Moreover, previous findings have demonstrated the effective use of smartphone applications for unsupervised digital memory assessment (e.g. Berron et al., 2022).

This body of work contributes towards research into concurrent multi-modal assessment, combining multiple types of data to improve diagnosis compared to any one measure. Of course, clinical decision-making is already based on triangulation of many factors and tests, but digital tools can provide large improvements in the amount of data collected per unit of time compared to, for example, a battery of neuropsychological tests. For example, this task combined spatial memory, memory monitoring, and eye-movements into one test. Only eye tracking was collected passively, but it is now feasible to combine VR with other physiological sensors including electrocardiography, electromyography, gait tracking, facial expression monitoring and electroencephalography. Although simply adding new modalities does not necessarily improve diagnosis, future research into new diagnostic tools can take advantage of these hardware capabilities, instead of focusing on one modality alone.

Appendix A

Clustering encoding scan-paths: a promising methodology

Abstract

A methodology was developed to analyse group distributions of similar scan-paths made during the encoding viewing period.

Methodology. First-person views of object configurations were clustered to find similar arrangements of objects. The MultiMatch shape measure was used as a distance metric for input into a clustering algorithm, resulting in clusters of similar scan-paths for each similar configuration of objects. The distribution of groups within scan-path clusters was compared to expected chance distributions.

Discussion and limitations. Unexpected noise introduced by task design precluded a robust analysis. This approach can be used to generate hypotheses around encoding strategies in future work.

A.1 Introduction

We can examine how participants viewed the same configuration of objects to provide a powerful means of comparing eye movements at encoding. Indeed, all participants were presented with the same set of configurations across all trials, without knowledge of the condition of the trial. This allowed direct comparison between how MCI+ participants viewed such a configuration in contrast to other groups. A potential outcome of this type of comparison is the identification of object configurations that were particularly discriminating of MCI+ participants. This could allow for a better understanding of neuropsychological impairment, and inform the development of future paradigms.

One approach could be to compare eye movements on a per-trial basis. However, comparing gaze patterns on each individual encoding phase assumes that each object configuration is unique. In fact, due to the pseudo-random generation of the object configurations, some were more similar than others, and can be grouped together. Here we encounter the first of two clustering opportunities in this chapter: clustering of object configurations. By grouping configurations together that form a similar spatial pattern, we can examine fixation sequences—or 'scan-paths'—between groups, while controlling for the shape formed by the objects.

At this point, one could examine viewing-level measures within configuration clusters, similar to results in previous chapters. However, we can explore the dataset further with a data-driven approach by leveraging techniques for quantifying the similarity between two scan-paths. For example, we have already examined the MultiMatch 'shape' metric for comparing the sequence of saccade vectors between two scan-paths (Dewhurst et al., 2012). A measure such as this can be used to group similar scan-paths together based on their spatial features. Here we encounter the titular clustering technique of this chapter: clustering the scan-paths themselves.

Using a similarity metric to generate clusters of scan-paths appears to be a

rare approach to analysing eye movements (although see Kumar et al., 2019). A technique such as this could be utilised to identify clusters of scan-paths that are more or less associated with a specific group, such as MCI+ participants. This in turn could be used to identify Alzheimer's-related fixation patterns that were either too subtle to detect with prior methods in this thesis, or were dependent on a particular spatial configuration of objects.

Unfortunately, unintended noise was introduced in the design of the task to preclude a full exploration of the utility of this approach in relation to the current study. Specifically, the rotation of each configuration itself was randomised relative to the viewer. This means that, although all participants viewed the same set of 54 configurations, each participant saw each configuration in a random one of four rotations, reducing the comparative power of this novel analysis. This drawback is somewhat counteracted by clustering similar configurations together, or rather clustering the participants' views of configurations. However, this resulted in some clusters of configurations with under-represented groups, and other clusters with multiple configurations from the same participant. The implications of this on the wider study are discussed later in this chapter. For now, this new methodology will be introduced as a promising methodology, rather than a results section for the thesis.

Accordingly, the remainder of this chapter will describe a methodology for clustering scan-paths by similarity and identifying imbalanced groups within those clusters. Exemplary scan-path clustering will be visualised, with statistical analyses to comparing group distributions within clusters.

A.2 Methods

A.2.1 Clustering object positions across configurations

Before clustering similar scan-paths themselves, object configurations were clustered by their spatial features. After some experimentation, the best results for this were attained by clustering configurations based on (a) object positions, (b) ordered vectors between object positions, and (c) the number of objects in the convex hull of the configuration. This latter feature was included because a handful of configurations had three objects in the convex hull, with the majority having four. The general procedure for clustering configurations went as follows:

1. Normalise the positions of configuration objects to the centroid (2D mean) of the configuration, oriented towards the participant's viewpoint based on their average head position during viewing.

2. Save the normalised object positions, the number of objects in their convex hull, and the vectors between normalised object positions ordered in the same direction (e.g. clockwise).

3. Define clustering parameters, depending on the specific clustering algorithm (see below).

4. Input object position, vector, and convex hull data as features into a clustering algorithm.

5. Plot results per cluster for visual inspection.

6. Iterate 2-4 based on visual inspection to avoid overly-liberal clustering parameters.

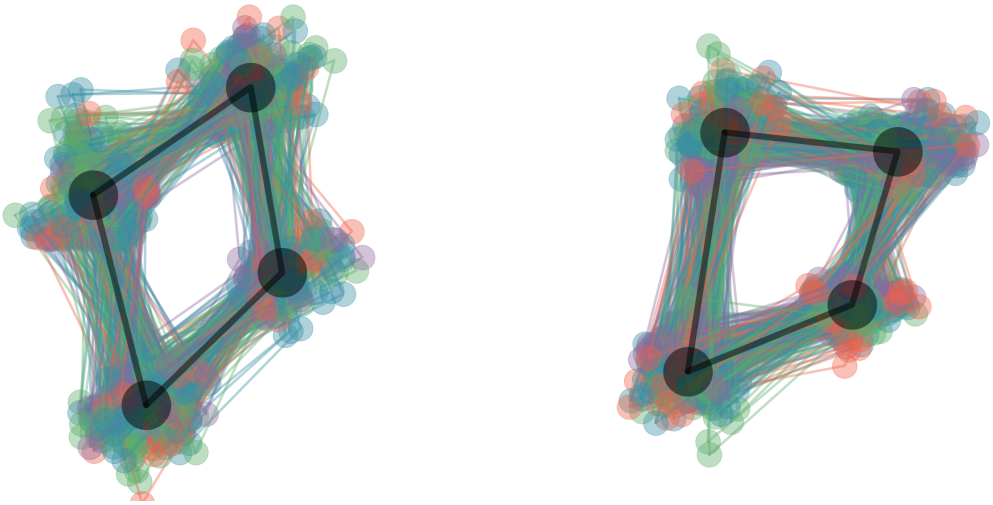


Figure A.1. Two examples of configuration clusters visualised by individual object positions per participant view (coloured dots) and averaged object positions (black dots). Lines between dots show the convex hull formed by the shape of the the object positions. Colours represent different participant groups consistent with previous sections: green = Younger, blue = Older, purple = MCI-, red = MCI+. Variation in configuration positions is introduced by difference in configurations themselves, but also from slight variation in head positions for each participant.

The needs of this clustering task are well met by agglomerative hierarchical clustering. This technique does not require a pre-defined number of clusters, which are unknown here, and a distance threshold can be applied to customise how liberal the clusters are (with a trade-off against the total number of clusters). The algorithm works by first treating each individual configuration as its own cluster, then using a distance-minimising algorithm to join, or 'agglomerate', clusters together. The Ward minimisation algorithm was used here, which minimises the sum of squared differences, similar to K-means clustering (Ward, 1963). Each cluster joins with the next nearest cluster until the distance threshold is reached, after which no new agglomeration occurs. A distance threshold of 3.0 was applied, which represents the Euclidean distance between each configuration's scaled features (their positions, vectors, and number of objects in the convex hull). Agglomerative clustering was applied using the scikit-learn

package in Python (Pedregosa et al., 2011).

A.2.2 Clustering scan-paths within configuration clusters

To group similar scan-paths together within configuration clusters, each participant's fixation centre-point was normalised to the mean object positions for that configuration cluster. This was achieved by translating each fixation position by the distance to each object's position, such that closer object positions were weighted greater than objects further away according to a Gaussian function

$$\text{weight} = \exp\left(-\frac{\text{distance}^2}{2 \cdot \sigma^2}\right)$$

with $\sigma = 0.08$ i.e. the mean radius of the AOI spheres (see sections 3.2.6 and 3.3.3).

Next, a similarity measure was used to define the 'distance' between scan-paths, instead of the Euclidean distance used for configuration clustering. The MultiMatch shape metric was used here, but any measure that quantifies the similarity between two fixation sequences can be used. A matrix of pair-wise distances between each scan-path was then entered into a clustering algorithm to identify scan-path groupings.

The final step was to relate the data-derived clusters to participant groups. Below is an exemplary workflow for statistically comparing the expected within-cluster group balanced to the observed distribution:

1. Pre-define the number of clusters equal to the number of groups.
2. Cluster scan-paths using an algorithm that can pre-define the number of clusters.
3. Test for the distribution of groups between clusters, for example by using a

chi-squared (χ^2) test of the null hypothesis that the distribution of groups is equal to the expected distribution. An alternative metric could be the Adjusted Rand Index (ARI; Hubert and Arabie, 1985), which is similar to an accuracy measure based on pairwise similarity (Rand Index) adjusted against chance clustering.

4. Examine the scan-paths of configuration clusters with significantly imbalanced group clustering. This was done via simple visual inspection here, but more systematic approaches might compare scan-path features such as the number of transitions between two specific AOIs.

The clustering method for step 2 above requires pre-definition of the number of clusters, which is a well-known characteristic of *K*-means clustering. However, *K*-means assumes round or spherical cluster shapes (Lloyd, 1982), which is unlikely to hold for these data because the 'features' being compared are sequences of vectors. A more robust alternative is spectral clustering (Ng et al., 2001), which applies a clustering technique such as *K*-means to a set of eigenvalues derived from the similarity matrix. Spectral clustering has been shown to perform better for non-isotropic cluster shapes, and requires fewer assumptions about the underlying data compared to *K*-means clustering (Ng et al., 2001).

For each configuration cluster, spectral clustering was used to group individual scan-paths into 4 clusters using the MultiMatch shape metric to create the distance matrix. A chi-squared test was then run for each configuration cluster to test whether the proportion of group counts in each scan-path cluster was over-representative of one or more groups. Configuration clusters with significantly imbalanced scan-path clusters were visually examined for scan-path differences.

A.3 Results

Configuration clustering yielded 38 distinct clusters. After adjusting for multiple comparisons using the Holm-Bonferroni procedure (Holm, 1979; $\alpha = .05$), no configurations were significantly imbalanced when clustering by the MultiMatch shape similarity metric. However, the configuration clusters with the two lowest *p*-values are shown in Figure A.2 as examples. One of these showed a potential over-representation of MCI+ participants, although this was not significant after multiple-comparison correction.

Figure A.3 shows examples of scan-paths for one imbalanced cluster, with group labels. Visual inspection of these plots can provide clues as to what features of the scan-paths led to the clusters. For example, in the top cluster (grey-blue) there seems to be a predominant anti-clockwise direction to the scan-path. Clusters two (yellow) and three (green) involve repeated saccades between the two right-hand AOIs, whereas the final (purple) cluster appears to involve saccades to and from the top right AOI.

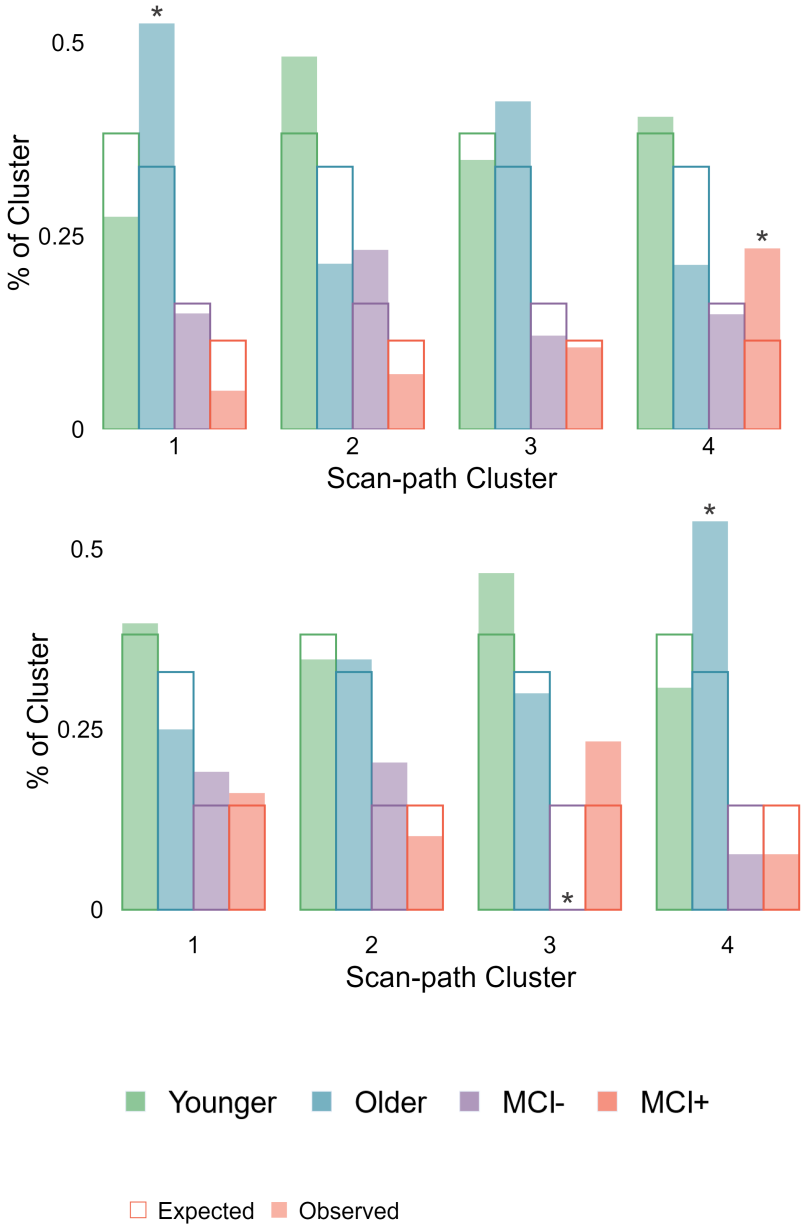


Figure A.2. Two configuration clusters with evidence for imbalanced groups across scan-path clusters, although not significantly after correction for multiple comparisons. Hollow coloured bars show expected group distribution per scan-path cluster. Filled coloured bars show actual % group prevalence within that cluster. Post hoc within-cluster significance was tested by examining adjusted standardised residuals per cell, * $p < .05$.

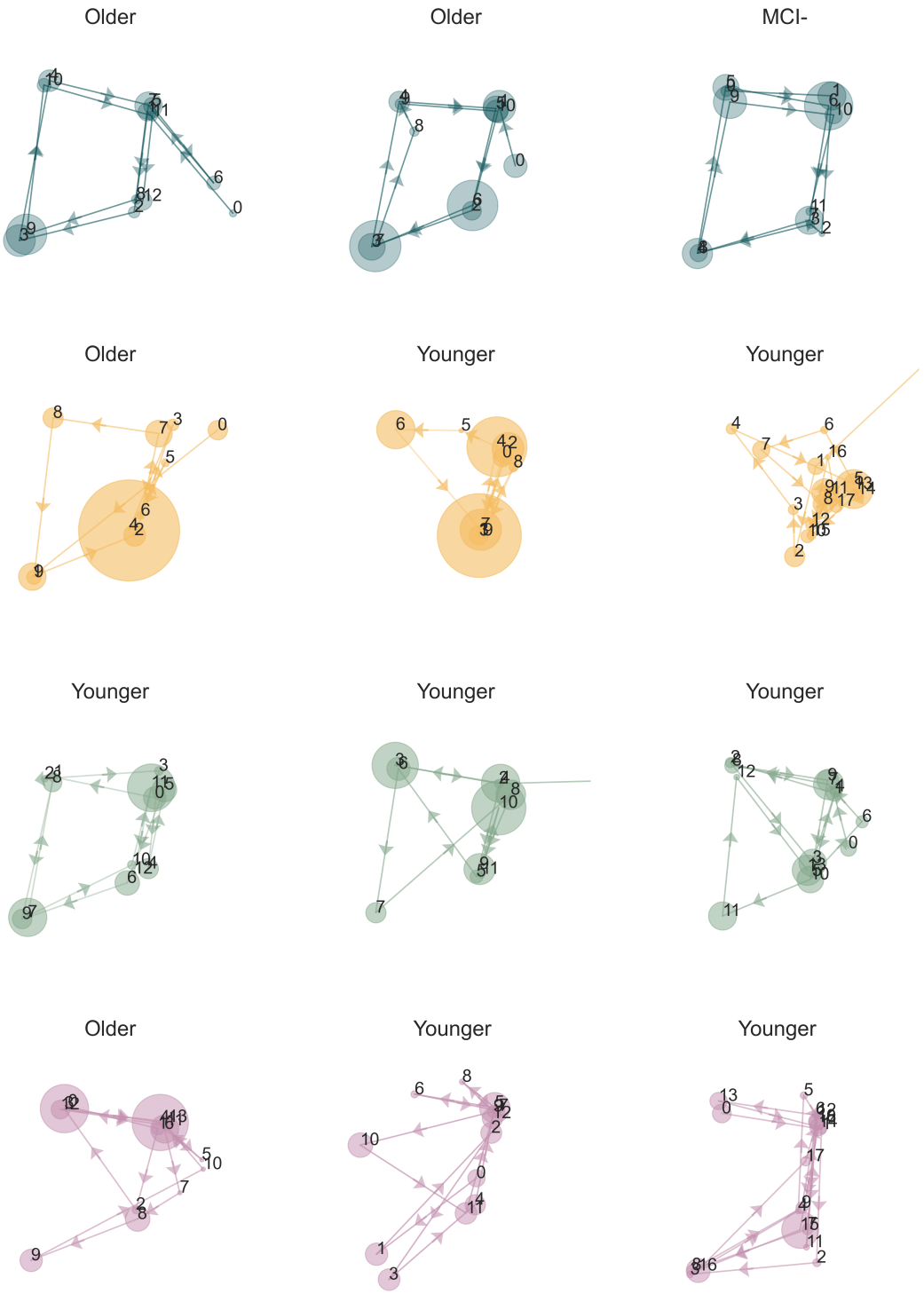


Figure A.3. Visualisations of fixation sequences scan-paths (fixation sequences) from one configuration cluster. Circles are external fixations (between AOIs, not within), lines are saccades between them. Circles are numbered by order of sequence, and sized by duration of fixation. Scan-path clusters are represented by different colours/rows. The participant's group is labelled above each scan-path. The balance of groups per row is *not* representative of the whole scan-path cluster.

A.4 Discussion

This chapter introduced a new methodology for clustering similar sequences of fixations based on pre-existing similarity measures. The approach was based on widely-available clustering algorithms combined with any measure of scan-path similarity, including published methods such as MultiMatch (Dewhurst et al., 2012). Although it cannot be used to draw conclusions about the current dataset, the technique provides a promising foundation for exploring encoding patterns in future work.

The scan-path clustering methodology can be used to identify different strategies used at encoding, or specific stimuli that discriminate healthy from impaired eye movements. For example, some participants showed a propensity to scan a specific configuration of four objects in a clockwise direction. Although this requires further investigation before concluding an effect, it is notable that several participants reported circular (i.e. clockwise or anti-clockwise) encoding strategies anecdotally. The advantage of the approach in this chapter is it provides a data-driven methodology for discovering potentially significant encoding eye movement strategies or patterns, which can then be used to test hypotheses with new data.

Of course, there are a number of limitations and caveats to this technique. One common issue with all statistical learning approaches, including clustering algorithms, is the effect of varying the hyperparameters, which govern how the algorithms learn or fit to the data. For clustering configurations this was the 'distance' threshold described earlier. A more extensive study of object configuration effects should make sure to test the effect of varying clustering distances. For clustering scan-paths, the hyper-parameter of the number of clusters was defined by the number of groups. However, pre-defining the number of clusters can lead to differing within-cluster variation as the algorithm forces outliers into clusters. An alternative approach would be to keep a distance threshold

and analyse scan-path clusters based on how distinct or similar they are to each other. An algorithm that would be well-suited to this approach is affinity propogation (Frey and Dueck, 2007), which would highlight outliers for further investigation.

As mentioned earlier in this chapter, task design issues limited the power of this methodology to a proof-of-concept. By clustering configuration views, the number of scan-paths per configuration varied and some configuration clusters contained multiple scan-paths from the same participant. This actually violates the assumption of independent observations for the χ^2 test. An approach that accounts for repeated measures, such as mixed-effects linear modelling, would be more appropriate for this specific use-case. However, I decided to keep the analysis simple and appropriate for the type of dataset it was intended for: that is, where independent scan-paths are clustered together only.

There is a further, study-wide implication here for having multiple participants per configuration cluster, which is that some participants essentially saw very similar configurations more than others. This has the potential to invalidate the counterbalancing of object configurations across conditions. However, given that configurations clustered into fewer groups than there were configurations, the spatial positioning within object configurations was not tightly controlled anyway. Furthermore, although configurations were clustered with a maximum distance threshold, there is a limit to the qualitative variation of four object positions given the arrangement constraints detailed in Section 2.2.

Interestingly, the effect of object configuration on memory performance in this and similar paradigms has rarely, to the best of my knowledge, been tested before, but tends to be counterbalanced or randomised to some extent (e.g. Heywood-Everett et al., 2022; Burgess et al., 2004). One group did find that varying whether an object shifted in relation to other objects mediated memory impairment in early AD (Ruggiero et al., 2020; Ruggiero et al., 2018). However, a future direction for this field of work could be to examine whether certain con-

figurations and object shifts are particularly difficult, especially when interacting with viewpoint shifts (see Section 7.4 for further discussion of this).

Appendix B

Supplementary Figures and Tables

B.1 Chapter 3 supplementary tables

Condition	N. Objects	df	t	p
Stay-Still	4	6	6.594	$<$0.001
Stay-Rotate	4	6	4.602	0.002
Walk-Still	4	6	7.778	$<$0.001
Walk-Rotate	4	6	7.706	$<$0.001
Teleport-Still	4	6	3.652	0.005
Teleport-Rotate	4	6	7.386	$<$0.001
Stay-Still	5	3	0.988	0.198
Stay-Rotate	5	3	3.900	0.015
Walk-Still	5	3	1.788	0.086
Walk-Rotate	5	3	0.598	0.296
Teleport-Still	5	3	0.850	0.229
Teleport-Rotate	5	3	0.899	0.217

Table B.1. Older feasibility participants' task performance compared to chance within each condition for four- and five-object groups. The higher *t* scores indicate greater difference from the chance score, which was 20% correct for five objects and 25% for four objects.

Condition	**N. Objects**	df	t	p
Stay-Still	4	6	14.201	$<$0.001
Stay-Rotate	4	6	4.647	0.002
Walk-Still	4	6	3.537	0.006
Walk-Rotate	4	6	5.851	0.001
Teleport-Still	4	6	4.240	0.003
Teleport-Rotate	4	6	4.648	0.002
Stay-Still	5	4	9.194	$<$0.001
Stay-Rotate	5	4	9.746	$<$0.001
Walk-Still	5	4	5.300	0.003
Walk-Rotate	5	4	7.004	0.001
Teleport-Still	5	4	3.849	0.009
Teleport-Rotate	5	4	3.077	0.019

Table B.2. Younger feasibility participants' task performance compared to chance within each condition for four- and five-object groups. The higher *t* scores indicate greater difference from the chance score, which was 20% correct for five objects and 25% for four objects.

B.2 Chapter 4 supplementary figures

B.2.1 Neuropsychological results

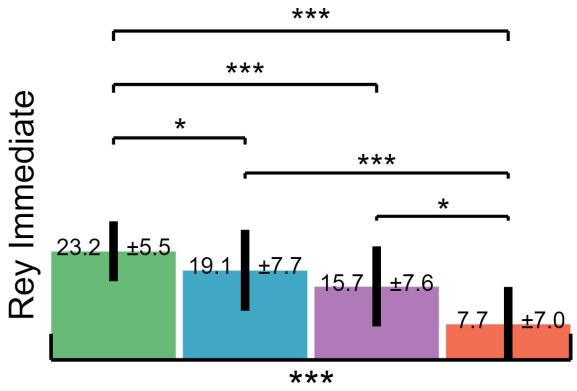


Figure B.1. Rey-Osterrieth Complex Figure Immediate Recall scores between groups. There was a significant main effect of group on Rey Delayed Recall $(F(3,\ 77) = 13.9,\ p < .001)$. Younger participants remembered significantly more of the Rey complex figure immediately after copying it than all other groups, and MCI+ participants scored lower than all other groups.

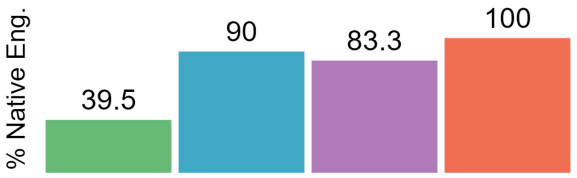


Figure B.2. Percentage of native english speakers in each participant group.

B.2.2 Eye movements on the moved object at retrieval

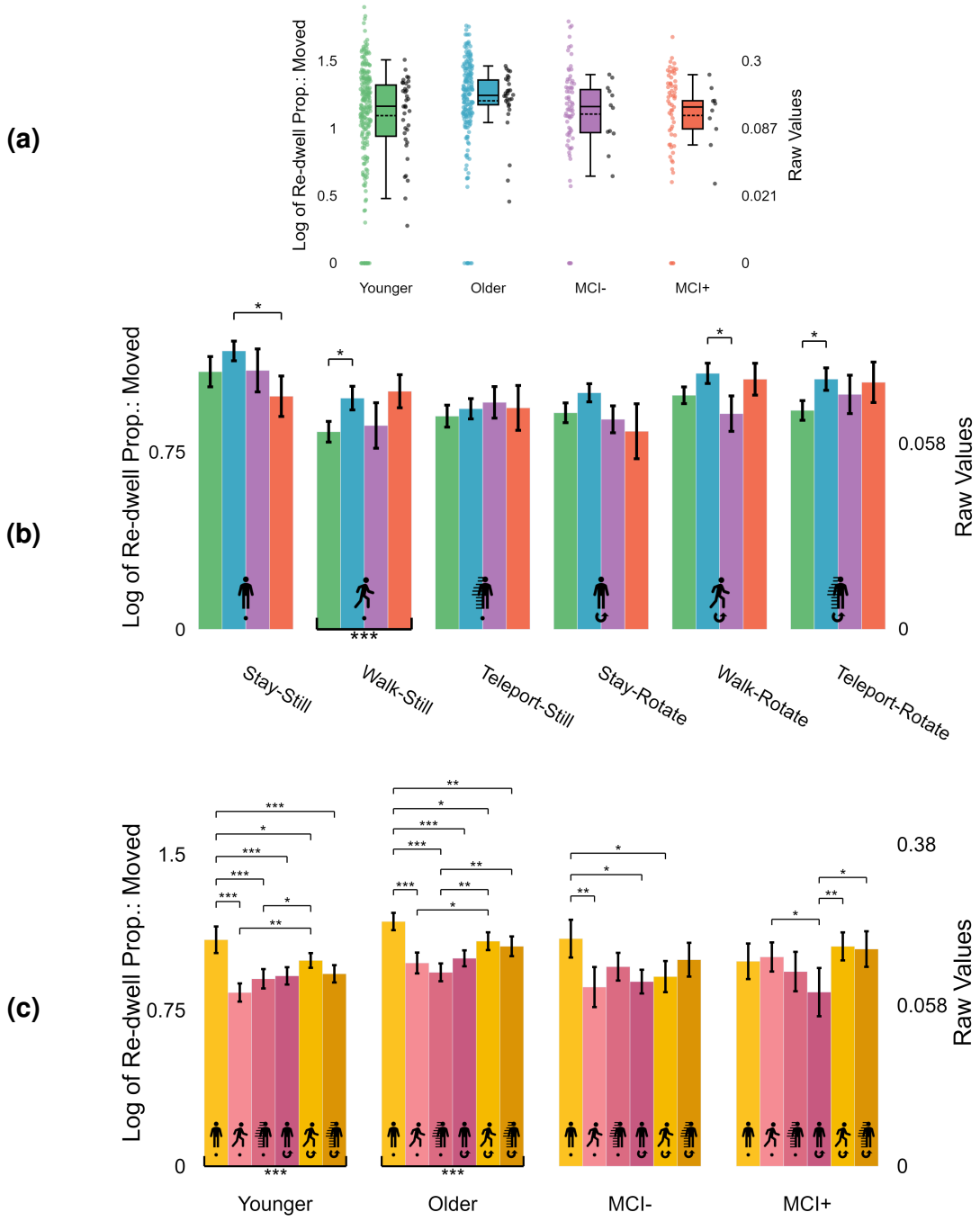


Figure B.3. Re-dwell on the moved object at retrieval. (a) group differences with all conditions combined (b) Within-condition group comparisons (c) Within-group condition comparisons. See figure 4.7 for more details of figure annotations.

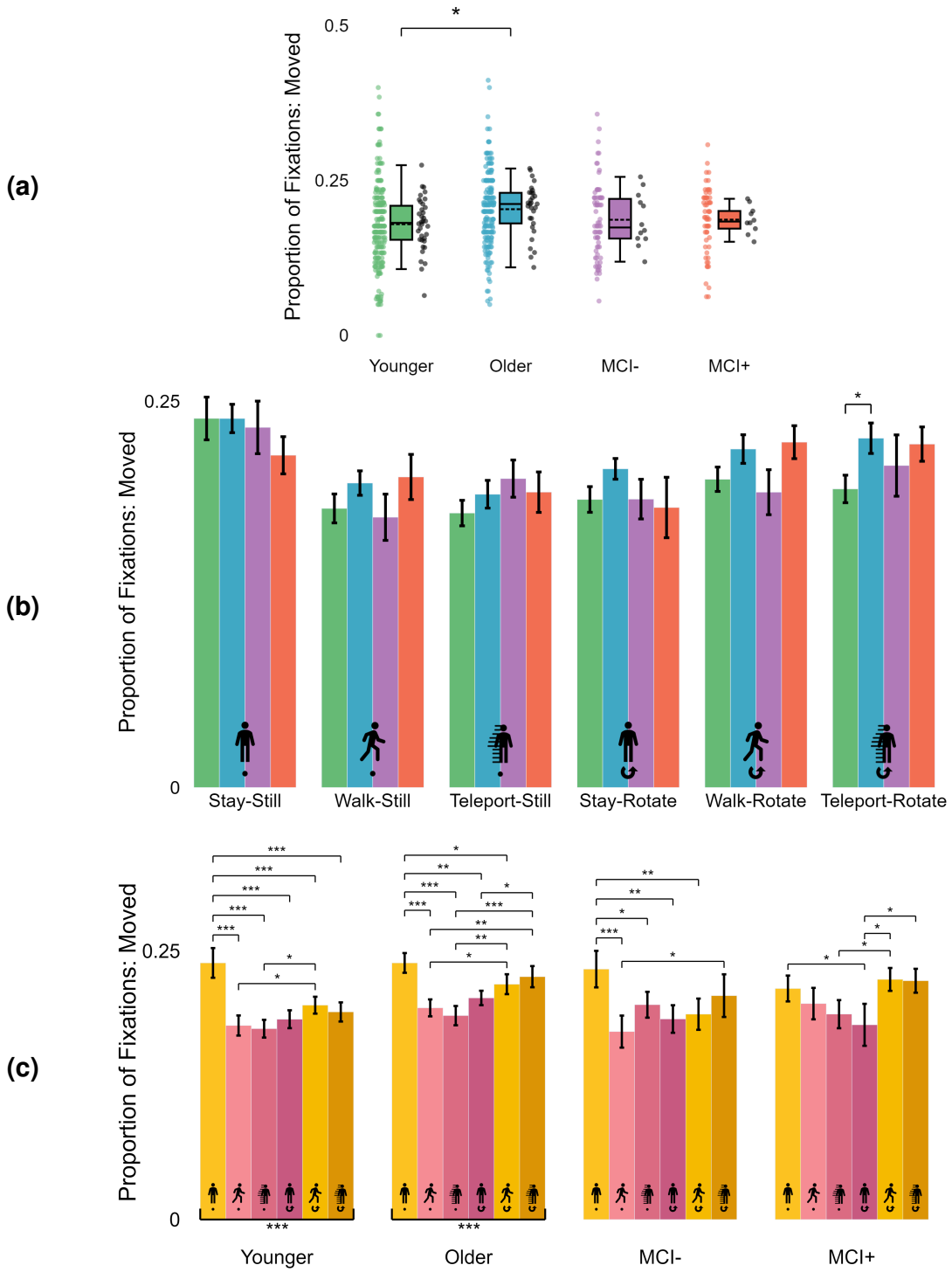


Figure B.4. Proportion of fixations on the moved object at retrieval. (a) group differences with all conditions combined (b) Within-condition group comparisons (c) Within-group condition comparisons. See figure 4.7 for more details of figure annotations.

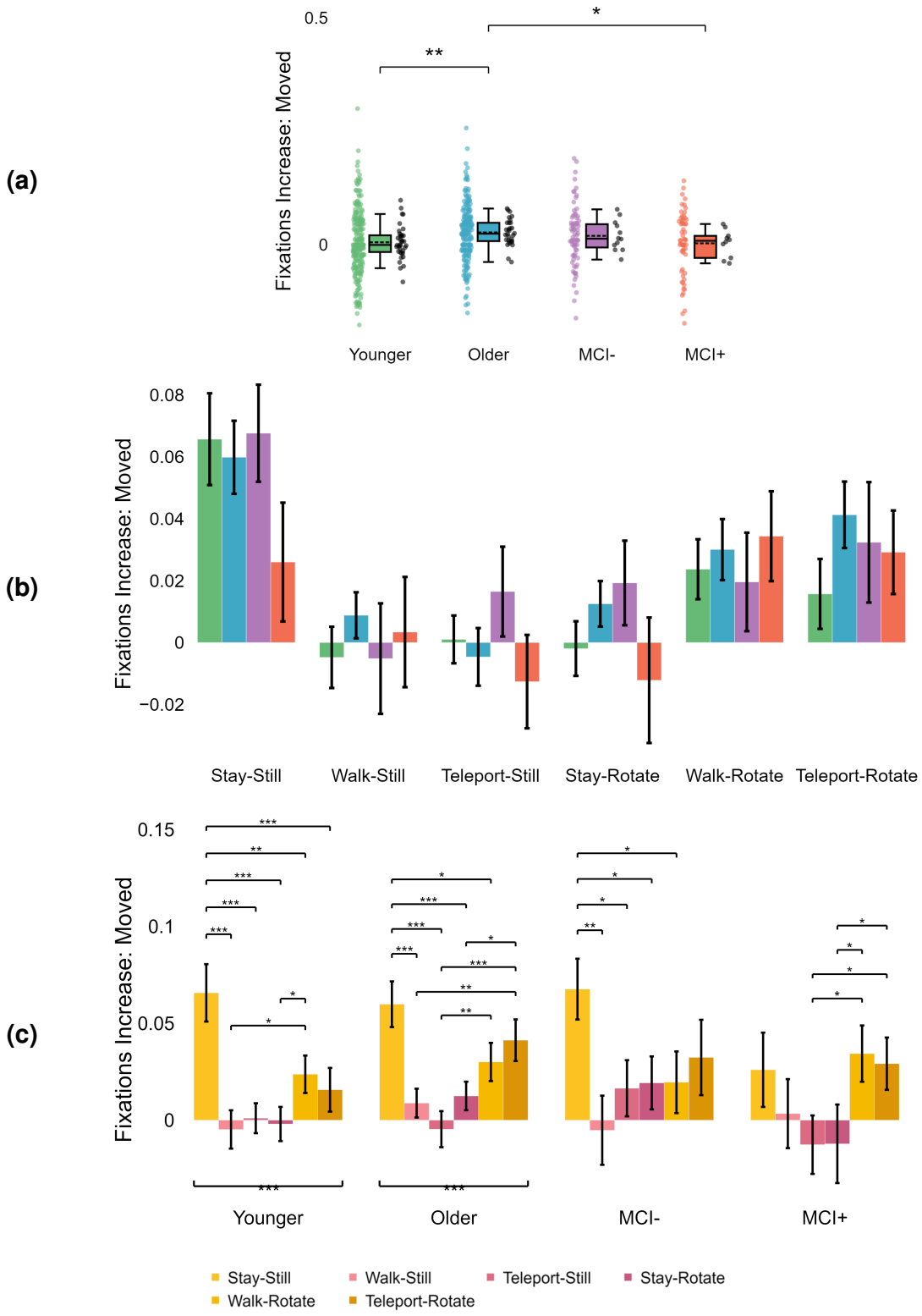


Figure B.5. Increase in proportion of fixations on the moved object from encoding to retrieval. (a) group differences with all conditions combined (b) Within-condition group comparisons (c) Within-group condition comparisons. See figure 4.7 for more details of figure annotations.

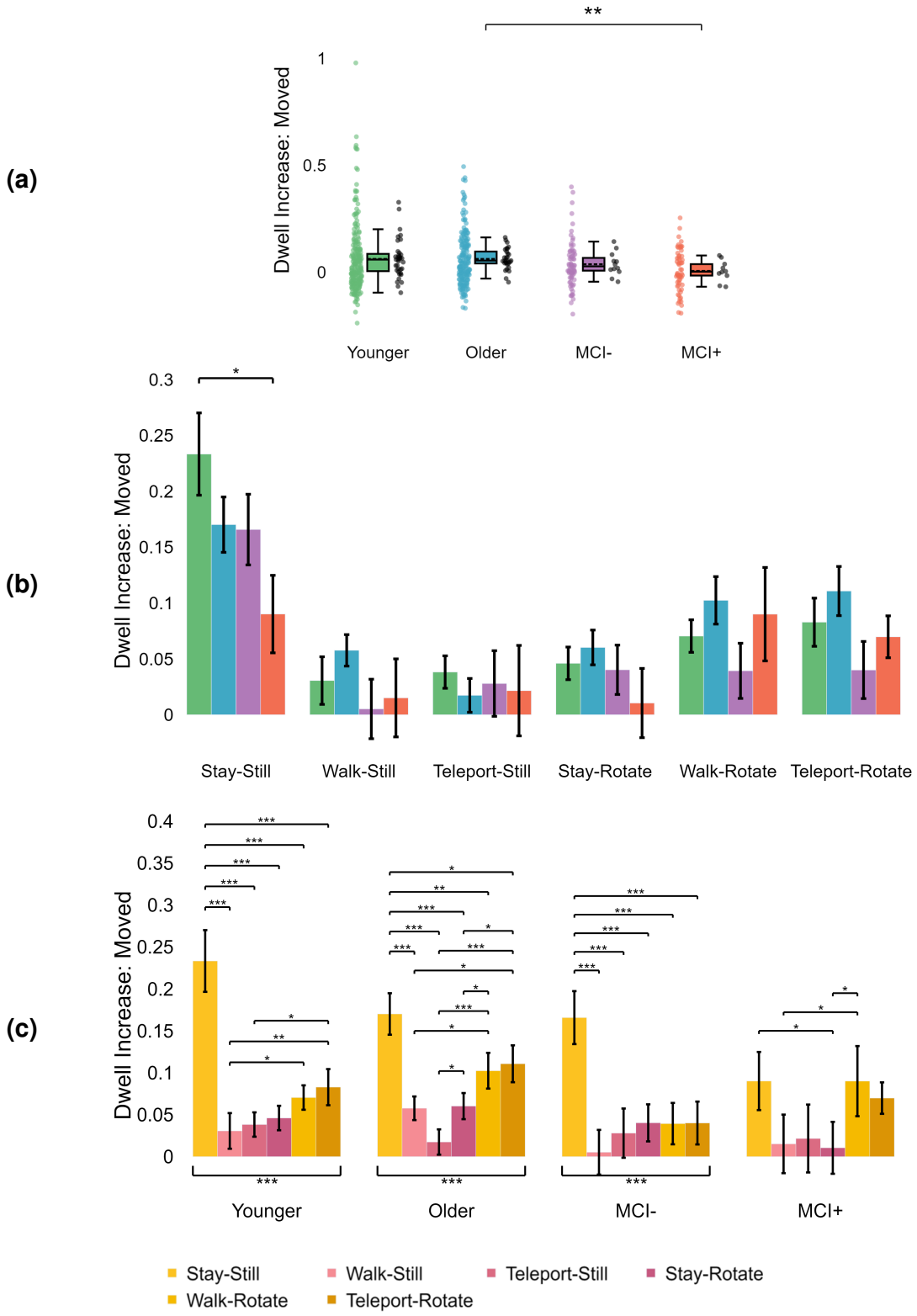


Figure B.6. Increase in dwell on the moved object from encoding to retrieval. (a) group differences with all conditions combined (b) Within-condition group comparisons (c) Within-group condition comparisons. See figure 4.7 for more details of figure annotations.

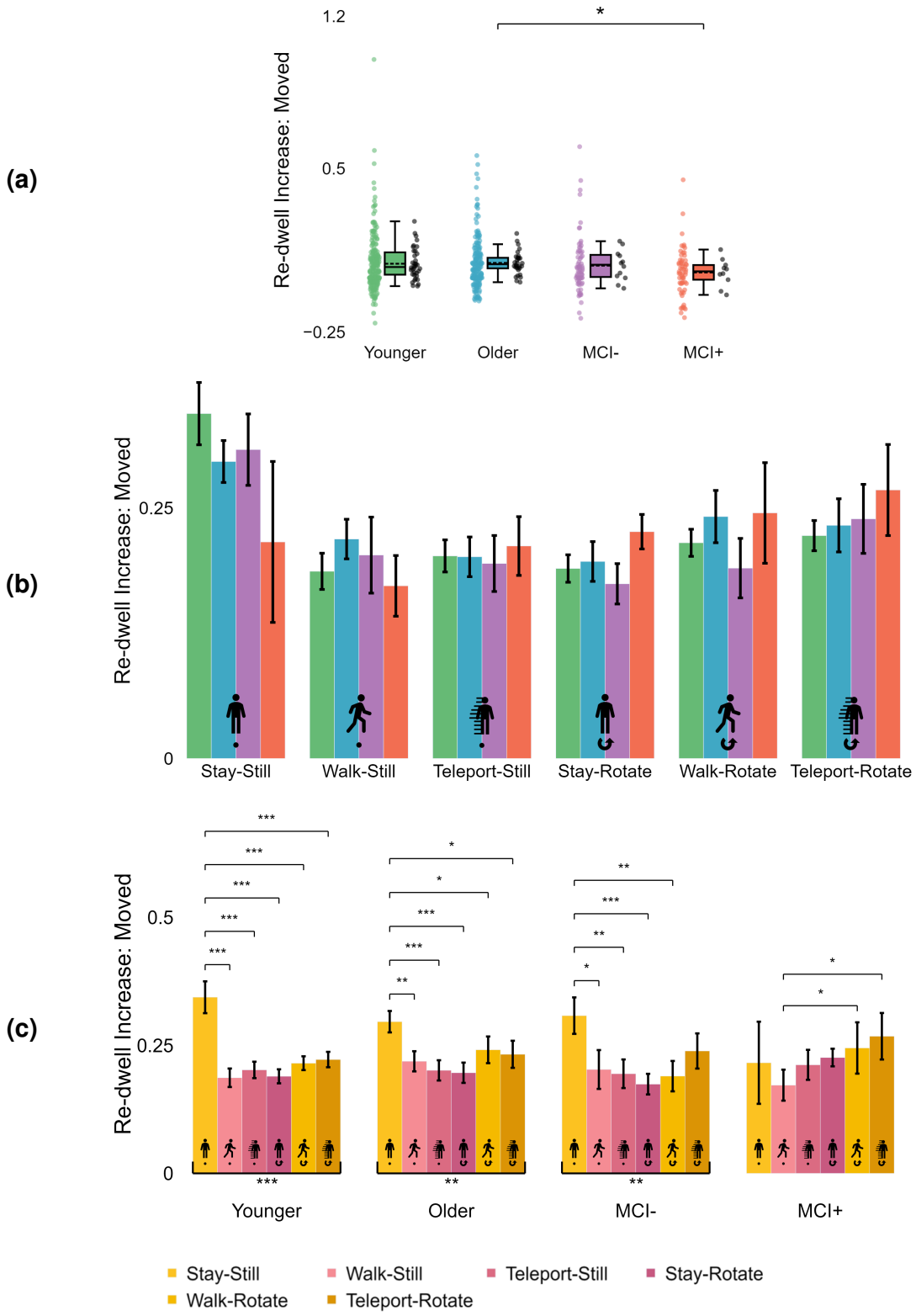


Figure B.7. Increase in re-dwell on the moved object from encoding to retrieval. (a) group differences with all conditions combined (b) Within-condition group comparisons (c) Within-group condition comparisons. See figure 4.7 for more details of figure annotations.

B.2.3 Eye movements on the stationary objects at retrieval

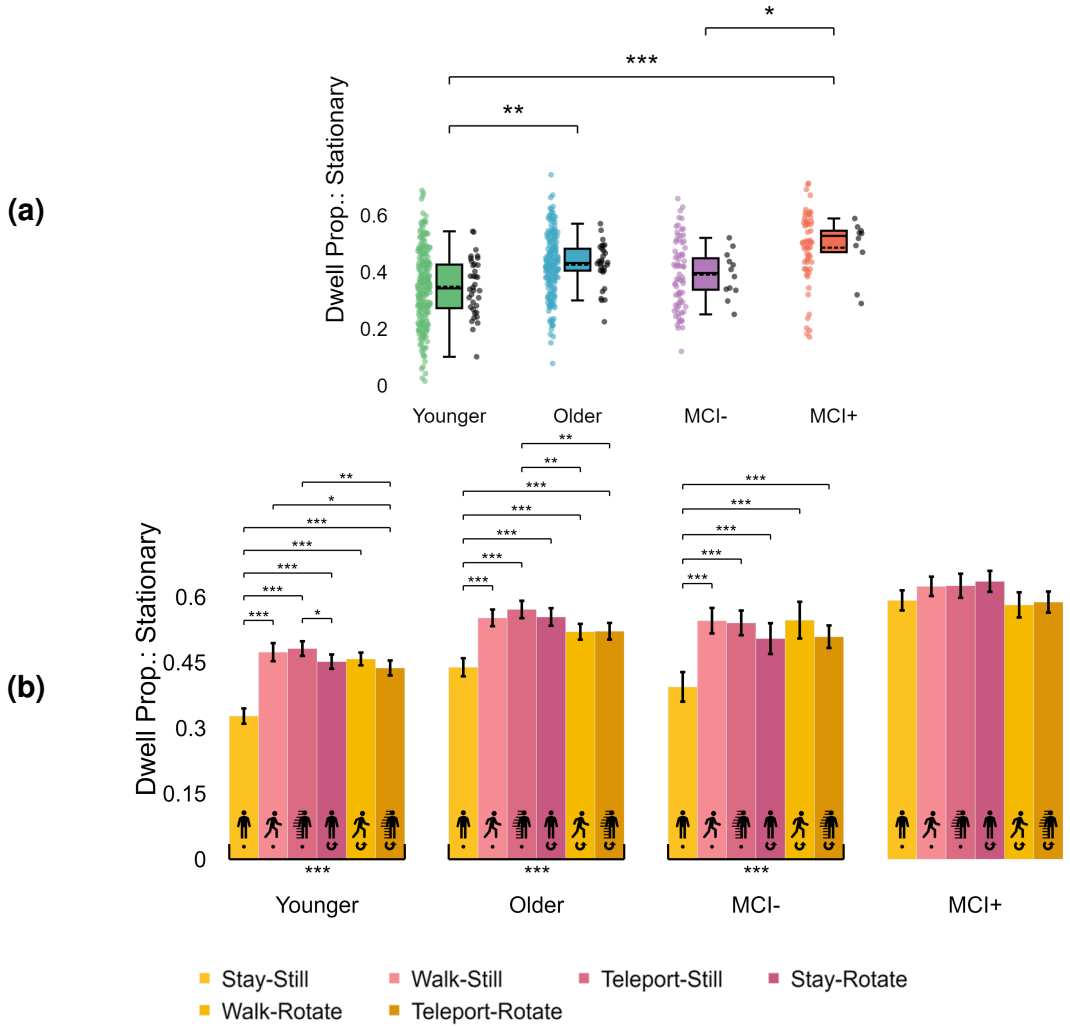


Figure B.8. Dwell on the stationary objects at retrieval (a) group differences with all conditions combined. (b) Within-group condition comparisons. For (b), a main effect of condition was found when including all groups ($F(5,\ 425) = 36.7,\ p < .001$). A significant interaction between group and condition was found when including only MCI groups ($F(5,\ 100) = 3.7,\ p = .004$), but not for healthy groups ($F(5,\ 325) = 0.9,\ p = .47$). Visual inspection of the chart suggests that the MCI+ group is the only showing no difference between conditions, whereas the other three groups have significantly reduced dwell proportion on the stationary objects in the Stay-Still condition at retrieval. See figure 4.7 for more details of figure annotations.

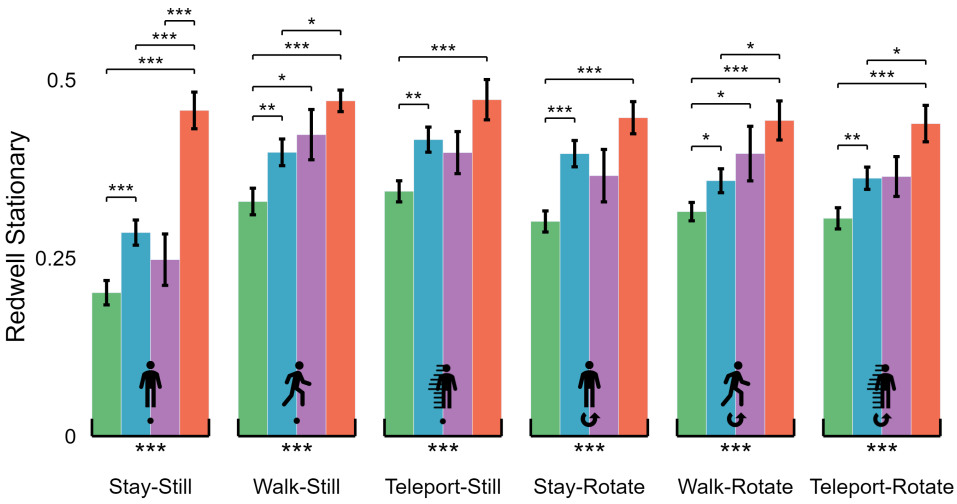


Figure B.9. Re-dwell on the stationary objects: within-condition group comparisons of dwell proportion from re-visits.

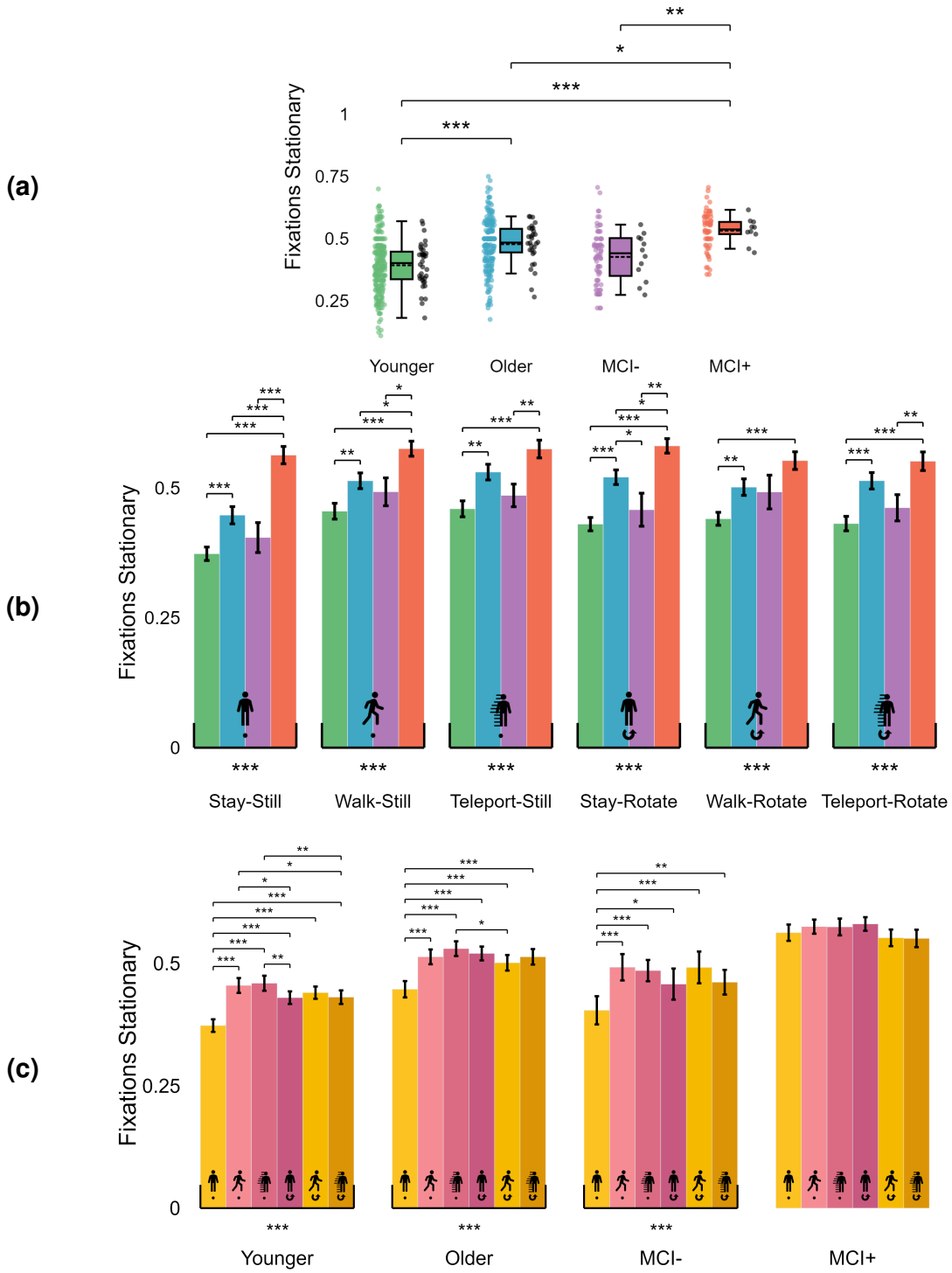


Figure B.10. Proportion of fixations on stationary objects at retrieval.

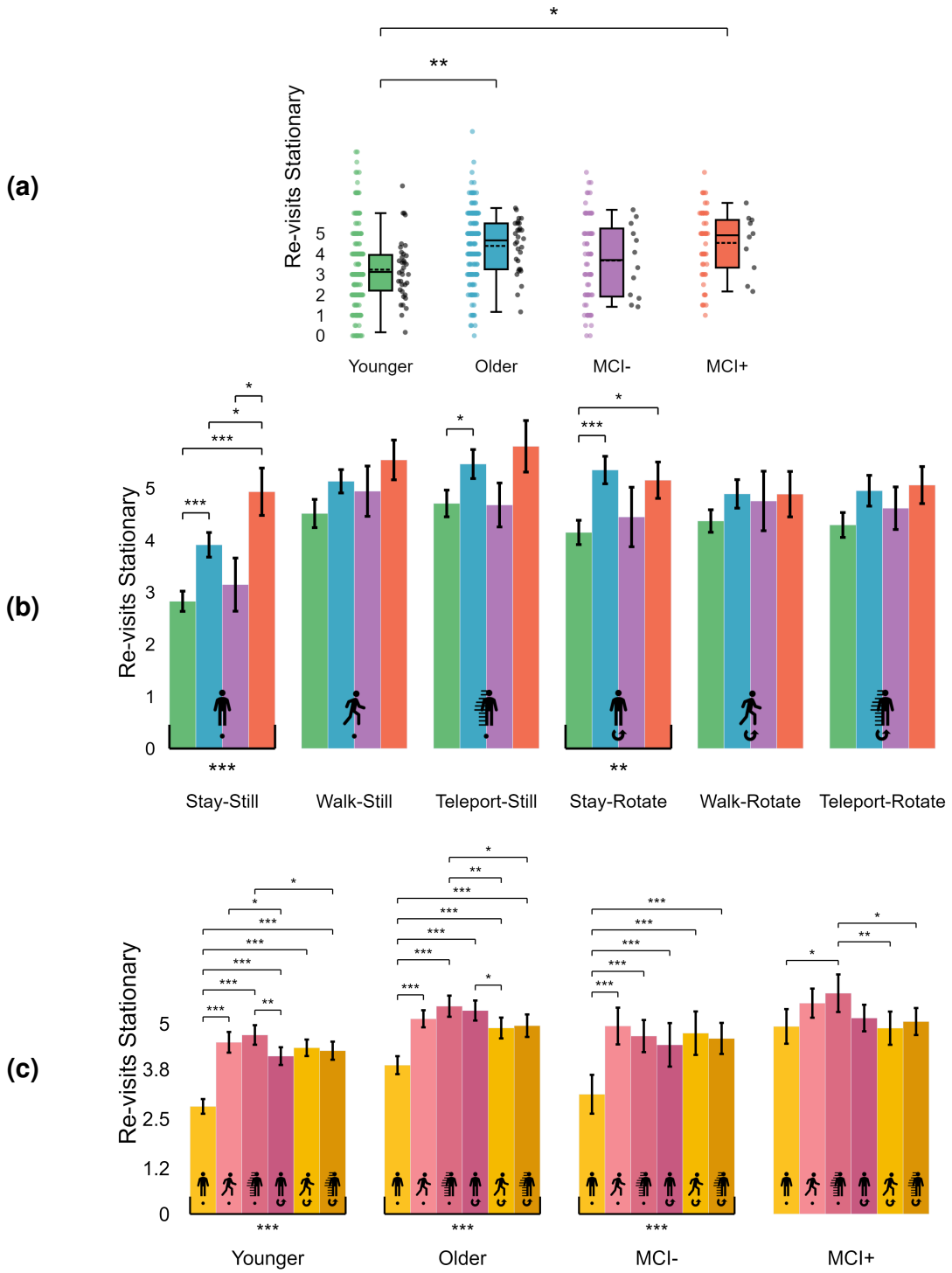


Figure B.11. Number of re-fixations on stationary objects at retrieval.

B.2.4 Eye movements on the table at retrieval

(a)

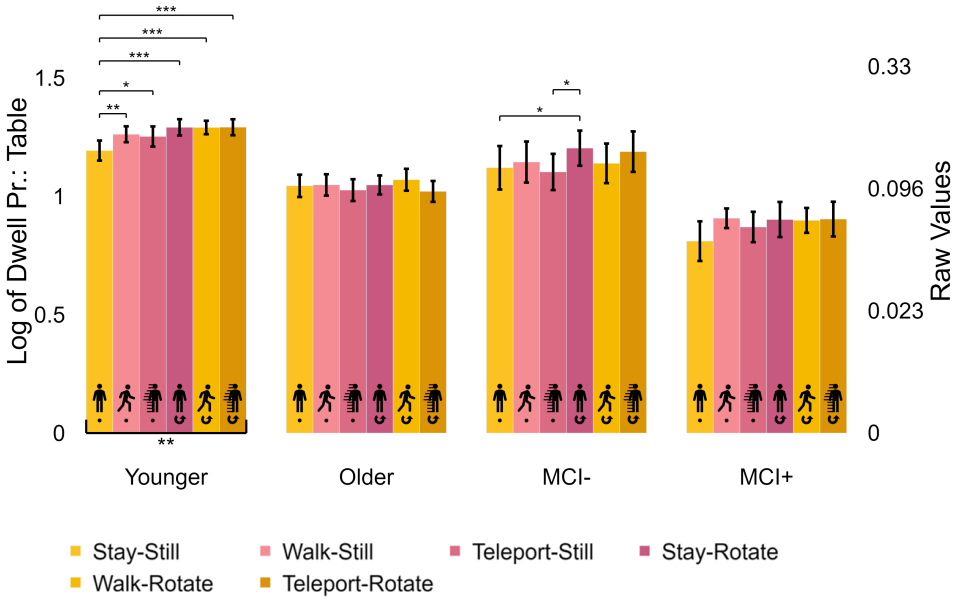


Figure B.12. Dwell on the table at retrieval within conditions. See figure 4.7 for more details of figure annotations.

B.2.5 Eye movements at encoding

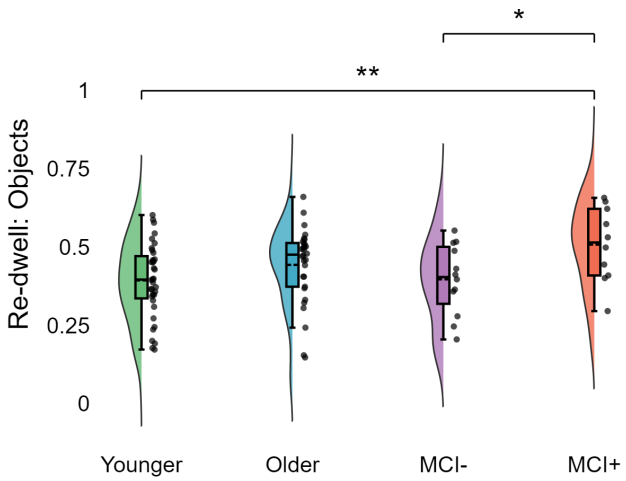


Figure B.13. Raincloud plots (Allen et al., 2021) of group differences in dwell proportions on configuration objects from revisits only, at encoding. There was a main effect of group on the re-dwell proportion on the objects $(F(3, 84) = 3.0, p = .04)$. The MCI+ group had significantly greater re-dwell on the objects compared to MCI- and Younger participants, but not MCI-.

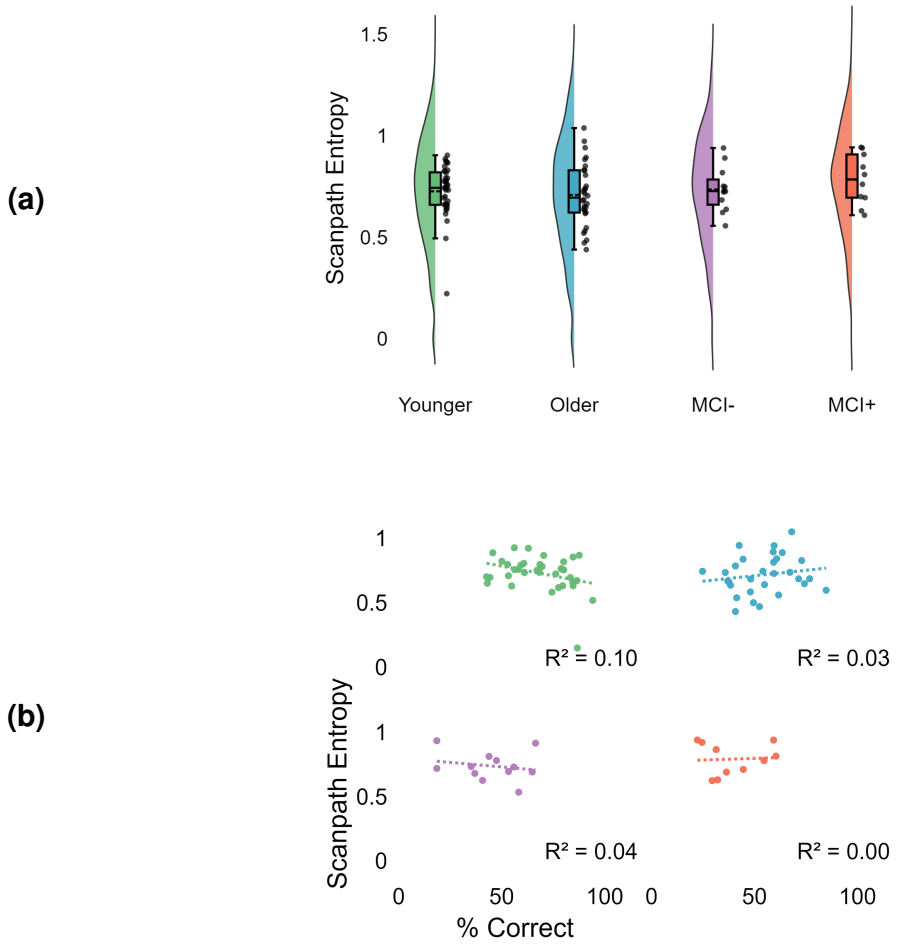


Figure B.14. (a) Raincloud plot of group differences in scan-path entropy at retrieval. (b) Scatter plots of scan-path entropy by percentage of correct trials, compared per group. Lines are linear models; solid lines are statistically significant, $p < .05$; dashed lines are not.

B.2.6 Post hoc eye metrics: moved object bias

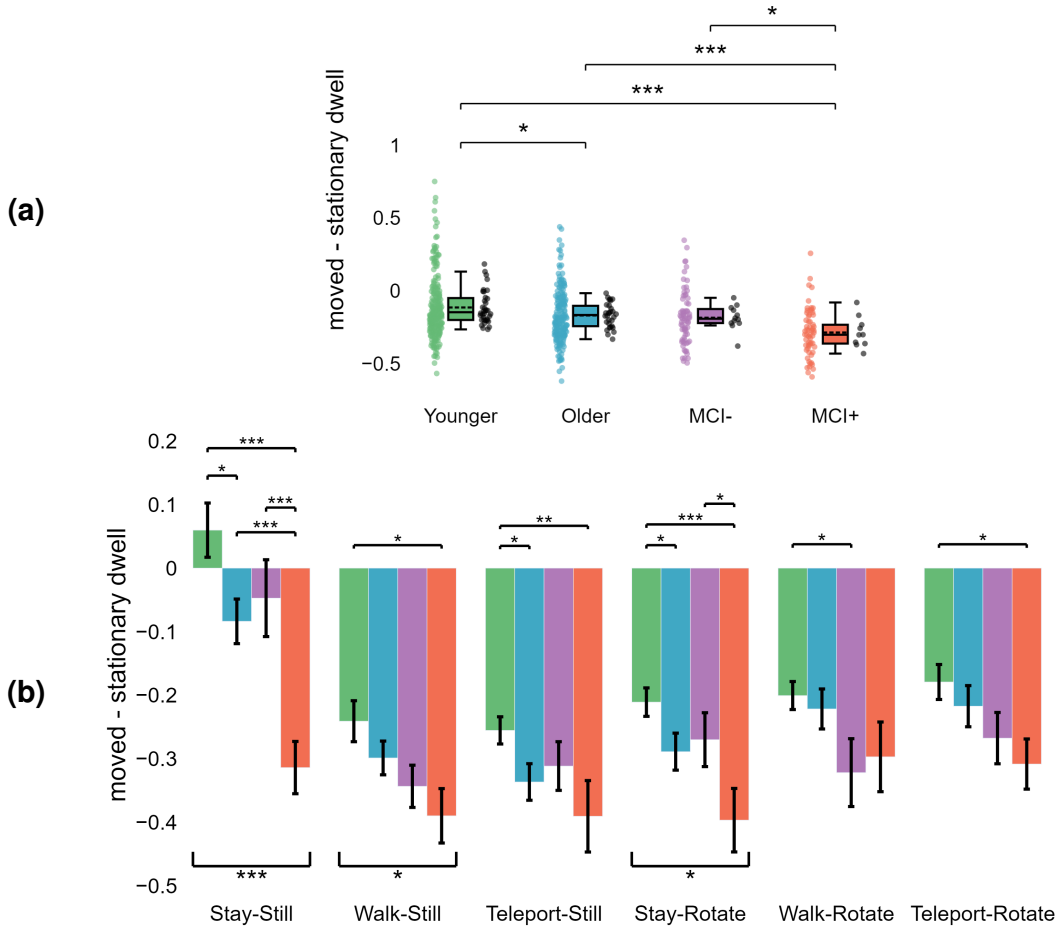


Figure B.15. Dwell proportion difference on the moved vs stationary objects at retrieval a.k.a. moved object bias. (a) Across all conditions between groups. There was a significant main effect of group ($F(3,\ 85) = 8.3,\ p < .001$) with pairwise differences between MCI+ and all other groups. This measure was calculated at the condition-level here. (b) Within-condition group comparisons. A positive bar indicates that average dwell proportion on the moved object was higher than for the other three stationary objects combined. A significant interaction between group and condition was found all groups combined ($F(15,\ 425) = 1.8,\ p = .03$), and for MCI groups only ($F(5,\ 100) = 3.5,\ p = .006$), but not for healthy groups ($F(5,\ 325) = 1.5,\ p = .20$). A significant pairwise reduction in moved object bias in MCI+ participants compared to all other groups was greatest in the Stay-Still condition (MCI+ vs Older $t(38) = -3.64,\ p = .001,\ d = -1.33$; MCI+ vs MCI- $t(20) = -3.75,\ p = .001,\ d = -1.61$).

B.2.7 Post hoc eye metrics: stationary avoidance and task performance

A strong main effect of trial correctness on stationary object avoidance was found when including all groups ($F(1, 85) = 255.1,\ p < .001,\ \eta_g^2 = .75$), only healthy groups ($F(1, 65) = 160.7,\ p < .001,\ \eta_g^2 = .71$), and only MCI groups ($F(1, 20) = 98.0,\ p < .001,\ \eta_g^2 = .83$). This suggests that all groups had a difference in stationary avoidance scores between correct and incorrect trials. Figure B.16a visualises a reduction in stationary avoidance in incorrect versus correct trials for all groups.

A small but significant interaction between group and trial correctness was found for healthy groups ($F(1, 65) = 11.6, p = .001, \eta_g^2 = .15$), suggesting a greater reduction in stationary avoidance in incorrect versus correct trials for either Older or Younger participants. Figure B.16a suggests the Older group may have a greater reduction in Stationary Avoidance for incorrect versus correct trials than the Younger group.

A significant interaction between condition and trial correctness in healthy groups ($F(5, 195) = 2.64, p = .025, \eta_g^2 = .14$) suggests that the effect of trial correctness on stationary object avoidance may differ depending on condition. Again, this seems to be driven by the difference between Stay-Still values and other conditions, which is greater for correct trials than incorrect trial in Younger, Older and MCI- groups (Figure B.16b, right).

For MCI groups, no significant interaction between group and correctness was found on stationary avoidance ($F(1, 20) = 0.68, p = 0.42$), suggesting that the reduction in this measure from correct to incorrect trials is similar for MCI- and MCI+ groups. However, a small but significant interaction between condition and correctness was found for these groups ($F(5, 80) = 2.93, p = .017, \eta_g^2 = .33$), indicating a variation in difference between conditions depending on trial correctness. This was likely driven by the MCI- group, whereas the MCI+ group did not show much difference between conditions in either correct or incorrect

trials (Figure B.16b, left). Accordingly, a three-way interaction effect was close to statistical significance , suggesting that this effect may exist but the study was under-powered to detect it.

Scatter plots in Figure B.16c show a significant association between mean stationary avoidance and task performance in the Older group, with positive but non-significant directional associations in all groups. This suggests a weak association between overall task performance and overall stationary avoidance at best.

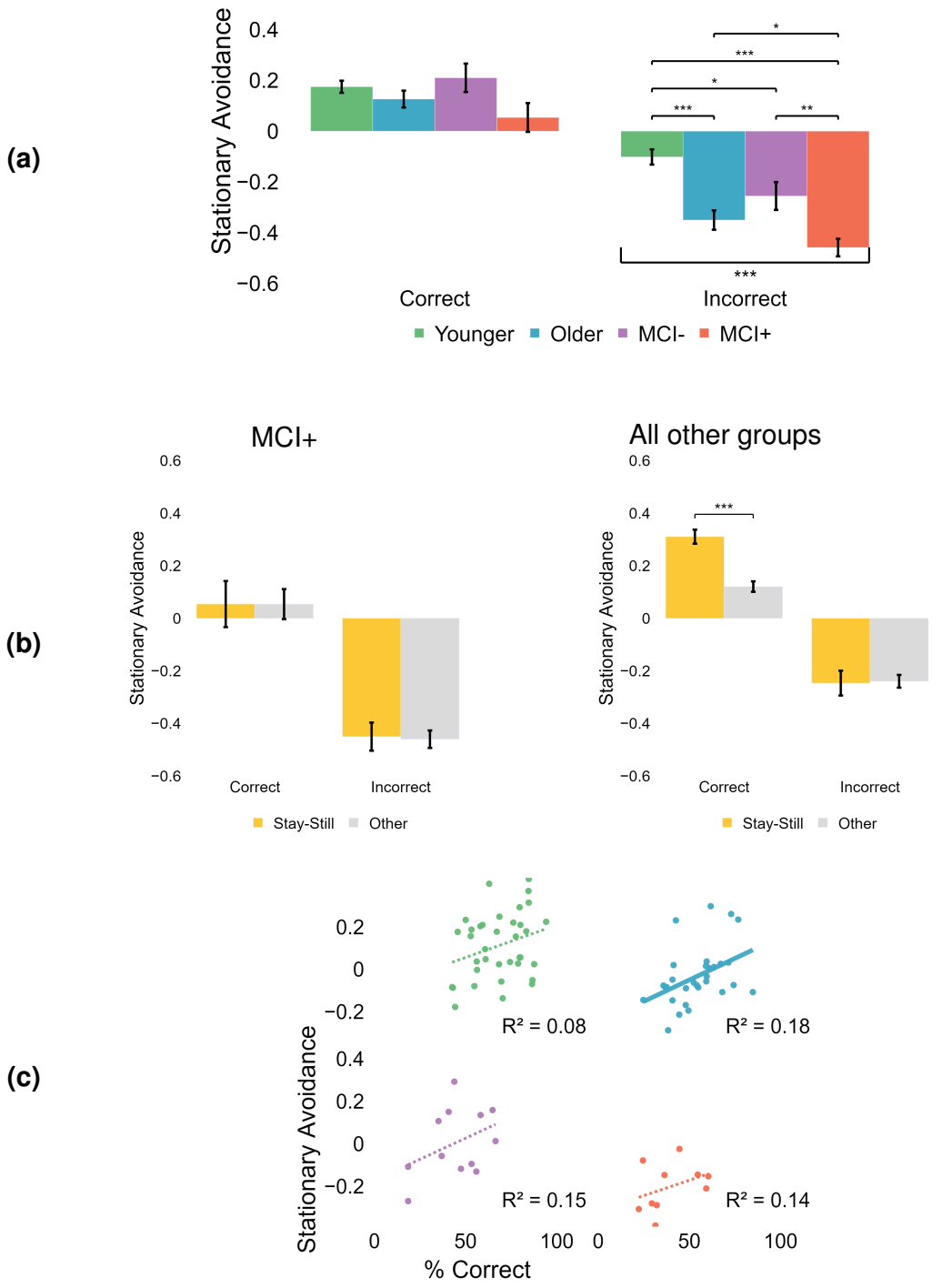


Figure B.16. Relationships between task performance and the stationary avoidance measure: (a) by group and trial correctness. Within-correctness significance is shown, between-correctness significance is not shown. (b) Stationary avoidance in Stay-Still compared to Other conditions, split by correctness for MCI+ participants (left) and Younger, Older and MCI- groups combined (right). (c) Scatter plots of stationary avoidance against percent of correct trials, split by group. See Figure 4.7 for more details of scatter plots.

B.3 Chapter 6 supplementary figures

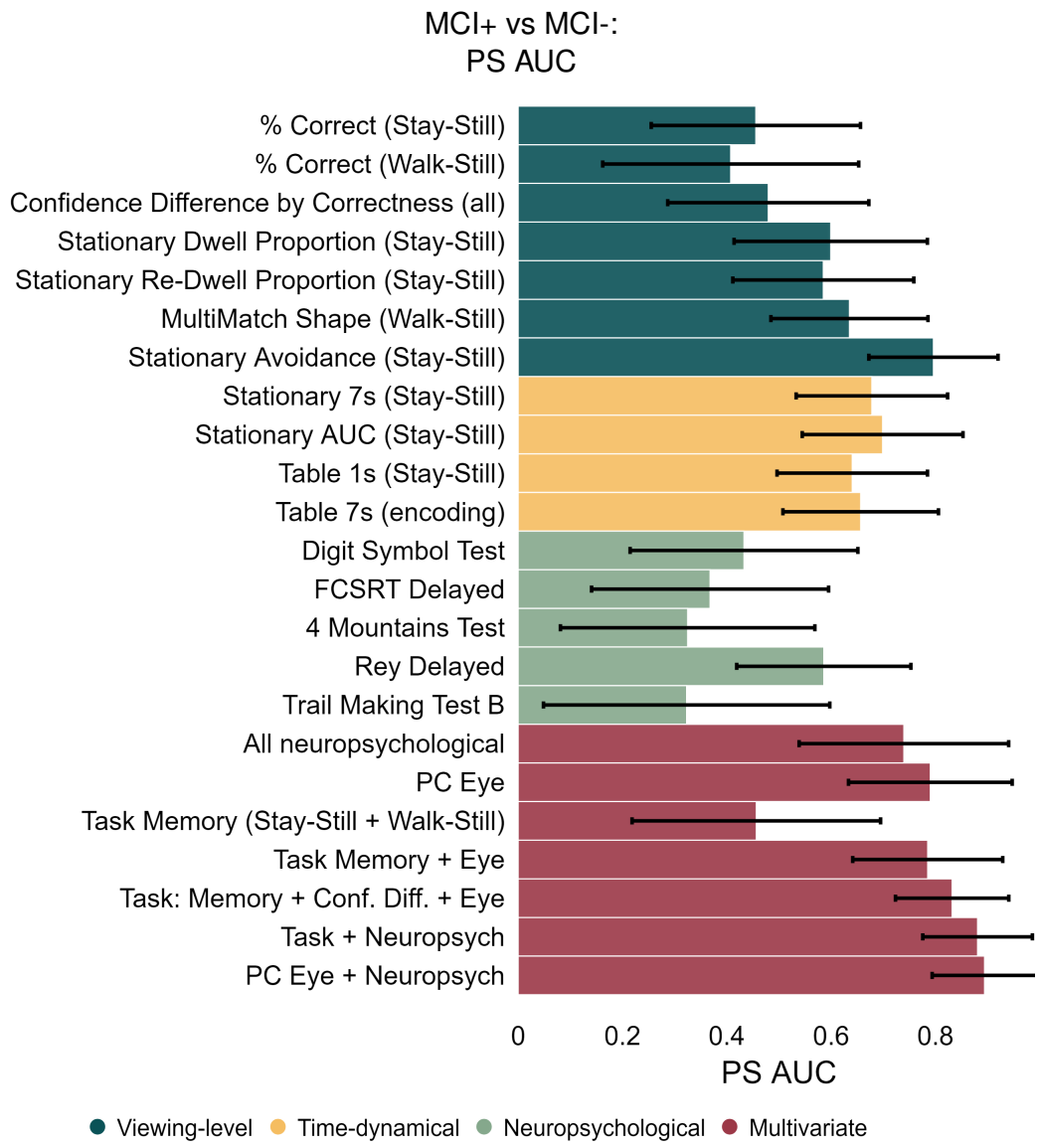


Figure B.17. Area Under Curves of Positive Predictive Value against Sensitivity aka precision-recall curve for classifying MCI+ from MCI- participants. Error bars are bootstrapped 95% confidence intervals from stratified bootstrapped cross-validation pipeline. Bars are coloured by the type of features the classifiers were trained on.

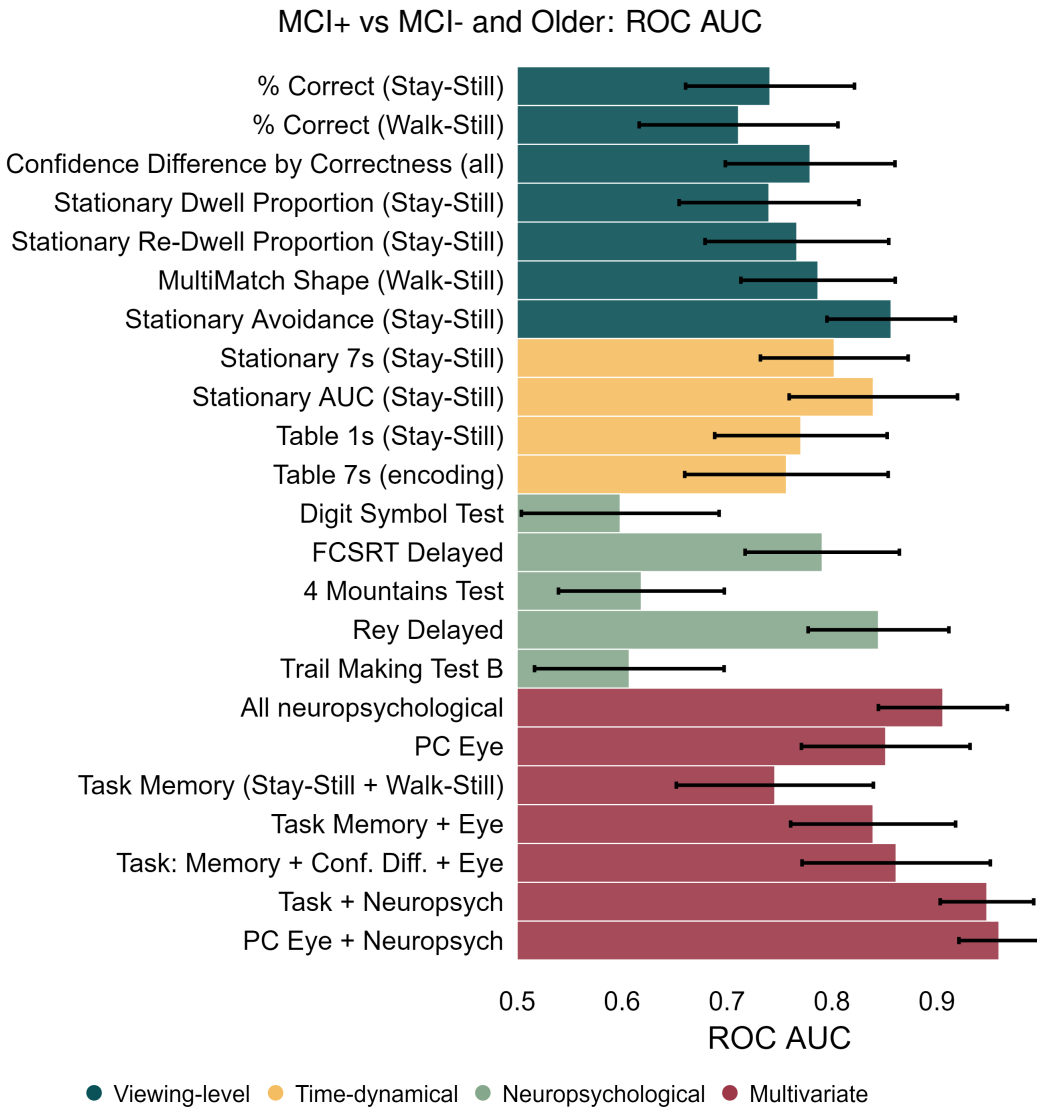


Figure B.18. Area Under the Receiver Operating Characteristics Curves by feature set, coloured by feature category, for classifying MCI+ participants from MCI- *and* Older participants. Error bars are bootstrapped 95% confidence intervals from stratified bootstrapped cross-validation pipeline. Bars are coloured by the type of features the classifiers were trained on.

References

Adhanom, I. B., Lee, S. C., Folmer, E., & MacNeilage, P. (2020). GazeMetrics: An Open-Source Tool for Measuring the Data Quality of HMD-based Eye Trackers. *ACM Symposium on Eye Tracking Research and Applications*, 1–5. https://doi.org/10.1145/3379156.3391374

Agrawal, Y., Smith, P. F., & Rosenberg, P. B. (2020). Vestibular impairment, cognitive decline and Alzheimer's disease: Balancing the evidence. *Aging & Mental Health*, *24*(5), 705–708. https://doi.org/10.1080/13607863.2019.1566813

Albert, M. S., DeKosky, S. T., Dickson, D., Dubois, B., Feldman, H. H., Fox, N. C., Gamst, A., Holtzman, D. M., Jagust, W. J., Petersen, R. C., Snyder, P. J., Carrillo, M. C., Thies, B., & Phelps, C. H. (2011). The diagnosis of mild cognitive impairment due to Alzheimer's disease: Recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimer's & Dementia*, *7*(3), 270–279. https://doi.org/10.1016/j.jalz.2011.03.008

Alfano, P. L., & Michel, G. F. (1990). Restricting the Field of View: Perceptual and Performance Effects. *Perceptual and Motor Skills*, *70*(1), 35–45. https://doi.org/10.2466/pms.1990.70.1.35

Allen, M., Poggiali, D., Whitaker, K., Marshall, T. R., van Langen, J., & Kievit, R. A. (2021). Raincloud plots: A multi-platform tool for robust data visualization. *Wellcome Open Research*, *4*, 63. https://doi.org/10.12688/wellcomeopenres.15191.2

Almkvist, O. (1996). Neuropsychological features of early Alzheimer's disease: Preclinical and clinical stages. *Acta Neurologica Scandinavica*, *94*(S165), 63–71. https://doi.org/10.1111/j.1600-0404.1996.tb05874.x

Almkvist, O., Basun, H., Bäckman, L., Herlitz, A., Lannfelt, L., Small, B., Viitanen, M., Wahlund, L. O., & Winblad, B. (1998). Mild cognitive impairment — an early stage of Alzheimer's disease? In H.-J. Gertz & T. Arendt (Eds.), *Alzheimer's Disease — From Basic Research to Clinical Applications* (pp. 21–29). Springer. https://doi.org/10.1007/978-3-7091-7508-8_3

Althnian, A., AlSaeed, D., Al-Baity, H., Samha, A., Dris, A. B., Alzakari, N., Abou Elwafa, A., & Kurdi, H. (2021). Impact of Dataset Size on Classification Performance: An Empirical Evaluation in the Medical Domain. *Applied Sciences*, *11*(2), 796. https://doi.org/10.3390/app11020796

Alzheimer, A. (1906). Uber einen eigenartigen schweren Er Krankungsprozeb der Hirnrinde. *Neurologisches Centralblatt*, *23*, 1129–1136.

Anderson, N. C., Anderson, F., Kingstone, A., & Bischof, W. F. (2015). A comparison of scanpath comparison methods. *Behavior Research Methods*, *47*(4), 1377–1392. https://doi.org/10.3758/s13428-014-0550-3

Andersson, R., Larsson, L., Holmqvist, K., Stridh, M., & Nyström, M. (2017). One algorithm to rule them all? An evaluation and discussion of ten eye movement event-detection algorithms. *Behavior Research Methods*, *49*(2), 616–637. https://doi.org/10.3758/s13428-016-0738-9

Andrews, J. S., Desai, U., Kirson, N. Y., Zichlin, M. L., Ball, D. E., & Matthews, B. R. (2019). Disease severity and minimal clinically important differences in clinical outcome assessments for Alzheimer's disease clinical trials. *Alzheimer's & Dementia: Translational Research & Clinical Interventions*, *5*(1), 354–363. https://doi.org/10.1016/j.trci.2019.06.005

Angel, L., Guerrerro-Sastoque, L., Bernardo, M., Vanneste, S., Isingrini, M., Bouazzaoui, B., Kachouri, H., Fay, S., & Taconnat, L. (2022). Metamemory mediates the protective effect of cognitive reserve on episodic memory during aging. *Acta Psychologica*, *228*, 103627. https://doi.org/10.1016/j.actpsy.2022.103627

Angelaki, D. E. (2014). How Optic Flow and Inertial Cues Improve Motion Perception. *Cold Spring Harbor Symposia on Quantitative Biology*, *79*, 141–148. https://doi.org/10.1101/sqb.2014.79.024638

Arboleda-Velasquez, J. F., Lopera, F., O'Hare, M., Delgado-Tirado, S., Marino, C., Chmielewska, N., Saez-Torres, K. L., Amarnani, D., Schultz, A. P., Sperling, R. A., Leyton-Cifuentes, D., Chen, K., Baena, A., Aguillon, D., Rios-Romenets, S., Giraldo, M., Guzmán-Vélez, E., Norton, D. J., Pardilla-Delgado, E., ... Quiroz, Y. T. (2019). Resistance to autosomal dominant Alzheimer's disease in an APOE3 Christchurch homozygote: A case report. *Nature Medicine*, *25*(11), 1680–1683. https://doi.org/10.1038/s41591-019-0611-3

Armstrong, R. A. (2019). Risk factors for Alzheimer's disease. *Folia Neuropathologica*, *57*(2), 87–105. https://doi.org/10.5114/fn.2019.85929

Bäckman, L., & Lipinska, B. (1993). Monitoring of general knowledge: Evidence for preservation in early Alzheimer's disease. *Neuropsychologia*, *31*(4), 335–345. https://doi.org/10.1016/0028-3932(93)90157-U

Bahill, A. T., Clark, M. R., & Stark, L. (1975). Dynamic overshoot in saccadic eye movements is caused by neurological control signal reversals. *Experimental Neurology*, *48*(1), 107–122. https://doi.org/10.1016/0014-4886(75)90226-5

Bartlett, C., Doyal, L., Ebrahim, S., Davey, P., Bachmann, M., Egger, M., & Dieppe, P. (2005). The causes and effects of socio-demographic exclusions from clinical trials. *Health Technology Assessment (Winchester, England)*, *9*(38), iii–iv, ix–x, 1–152. https://doi.org/10.3310/hta9380

Bature, F., Guinn, B.-a., Pang, D., & Pappas, Y. (2017). Signs and symptoms preceding the diagnosis of Alzheimer's disease: A systematic scoping review of literature from 1937 to 2016. *BMJ Open*, *7*(8), e015746. https://doi.org/10.1136/bmjopen-2016-015746

Bécu, M., Sheynikhovich, D., Tatur, G., Agathos, C. P., Bologna, L. L., Sahel, J.-A., & Arleo, A. (2019). Age-related preference for geometric spatial cues during real-world navigation. *Nature Human Behaviour*. https://doi.org/10.1038/s41562-019-0718-z

Belleville, S., Chertkow, H., & Gauthier, S. (2007). Working memory and control of attention in persons with Alzheimer's disease and mild cognitive impairment. *Neuropsychology*, *21*(4), 458–469. https://doi.org/10.1037/0894-4105.21.4.458

Bellman, R. (1966). Dynamic Programming. *Science*, *153*(3731), 34–37. https://doi.org/10.1126/science.153.3731.34

Beltrán, J., García-Vázquez, M. S., Benois-Pineau, J., Gutierrez-Robledo, L. M., & Dartigues, J.-F. (2018). Computational Techniques for Eye Movements Analysis towards Supporting Early Diagnosis of Alzheimer's Disease: A Review. *Computational and Mathematical Methods in Medicine*, *2018*, e2676409. https://doi.org/10.1155/2018/2676409

Berron, D., Ziegler, G., Vieweg, P., Billette, O., Güsten, J., Grande, X., Heneka, M. T., Schneider, A., Teipel, S., Jessen, F., Wagner, M., & Düzel, E. (2022). Feasibility of Digital Memory Assessments in an Unsupervised and Remote Study Setting. *Frontiers in Digital Health*, *4*. Retrieved November 23, 2023, from https://www.frontiersin.org/articles/10.3389/fdgth.2022.892997

Bhandari, J., MacNeilage, P., & Folmer, E. (2018). Teleportation without Spatial Disorientation Using Optical Flow Cues.

Bhopal, R. (2008). Ethical Issues in Health Research on Ethnic Minority Populations: Focusing on Inclusion and Exclusion. *Research Ethics*, *4*(1), 15–19. https://doi.org/10.1177/174701610800400104

Bicanski, A., & Burgess, N. (2019). A Computational Model of Visual Recognition Memory via Grid Cells. *Current Biology*, *29*(6), 979–990.e4. https://doi.org/10.1016/j.cub.2019.01.077

Bicanski, A., & Burgess, N. (2020). Neuronal vector coding in spatial cognition. *Nature Reviews Neuroscience*, *21*(9), 453–470. https://doi.org/10.1038/s41583-020-0336-9

Bigler, E. D., Rosa, L., Schultz, F., Hall, S., & Harris, J. (1989). Rey-Auditory Verbal Learning and Rey-Osterrieth Complex Figure Design performance in Alzheimer's disease and closed head injury. *Journal of Clinical Psychology*, *45*(2), 277–280. https://doi.org/10.1002/1097-4679(198903)45:2⟨277::AID-JCLP2270450216⟩3.0.CO;2-M

Biondi, J., Fernandez, G., Castro, S., & Agamennoni, O. (2018). Eye movement behavior identification for Alzheimer's disease diagnosis. *Journal of Integrative Neuroscience*, *17*(4), 349–354. https://doi.org/10.31083/j.jin.2018.04.0416

Bird, C. M., & Burgess, N. (2008). The hippocampus and memory: Insights from spatial processing. *Nature Reviews Neuroscience*, *9*(3), 182–194. https://doi.org/10.1038/nrn2335

Bird, C. M., Shallice, T., & Cipolotti, L. (2007). Fractionation of memory in medial temporal lobe amnesia. *Neuropsychologia*, *45*(6), 1160–1171. https://doi.org/10.1016/j.neuropsychologia.2006.10.011

Bird, C. M., Vargha-Khadem, F., & Burgess, N. (2008). Impaired memory for scenes but not faces in developmental hippocampal amnesia: A case study. *Neuropsychologia*, *46*(4), 1050–1059. https://doi.org/10.1016/j.neuropsychologia.2007.11.007

Bloom, G. S. (2014). Amyloid-beta and Tau: The Trigger and Bullet in Alzheimer Disease Pathogenesis. *JAMA Neurology*, *71*(4), 505–508. https://doi.org/10.1001/jamaneurol.2013.5847

Bogartz, R. S., & Staub, A. (2012). Gaze step distributions reflect fixations and saccades: A comment on Stephen and Mirman (2010). *Cognition*, *123*(2), 325–334. https://doi.org/10.1016/j.cognition.2011.12.014

Boletsis, C., & Chasanidou, D. (2022). A Typology of Virtual Reality Locomotion Techniques. *Multimodal Technologies and Interaction*, *6*(9), 72. https://doi.org/10.3390/mti6090072

Bowen, D. J., Kreuter, M., Spring, B., Cofta-Woerpel, L., Linnan, L., Weiner, D., Bakken, S., Kaplan, C. P., Squiers, L., Fabrizio, C., & Fernandez, M. (2009). How We Design Feasibility Studies. *American Journal of Preventive Medicine*, *36*(5), 452–457. https://doi.org/10.1016/j.amepre.2009.02.002

Braak, H., & Braak, E. (1991). Neuropathological stageing of Alzheimer-related changes. *Acta Neuropathologica*, *82*(4), 239–259. https://doi.org/10.1007/BF00308809

Brookmeyer, R., & Abdalla, N. (2018). Estimation of lifetime risks of Alzheimer's disease dementia using biomarkers for preclinical disease. *Alzheimer's & Dementia*, *14*(8), 981–988. https://doi.org/10.1016/j.jalz.2018.03.005

Buckner, R. L. (2004). Memory and Executive Function in Aging and AD: Multiple Factors that Cause Decline and Reserve Factors that Compensate. *Neuron*, *44*(1), 195–208. https://doi.org/10.1016/j.neuron.2004.09.006

Bueno, A. P. A., Sato, J. R., & Hornberger, M. (2019). Eye tracking – The overlooked method to measure cognition in neurodegeneration? *Neuropsychologia*, *133*, 107191. https://doi.org/10.1016/j.neuropsychologia.2019.107191

Burgess, N. (2006). Spatial memory: How egocentric and allocentric combine. *Trends in Cognitive Sciences*, *10*(12), 551–557. https://doi.org/10.1016/j.tics.2006.10.005

Burgess, N. (2008). Spatial Cognition and the Brain. *Annals of the New York Academy of Sciences*, *1124*(1), 77–97. https://doi.org/10.1196/annals.1440.002

Burgess, N., Maguire, E. A., & O'Keefe, J. (2002). The Human Hippocampus and Spatial and Episodic Memory. *Neuron*, *35*(4), 625–641. https://doi.org/10.1016/S0896-6273(02)00830-9

Burgess, N., Spiers, H. J., & Paleologou, E. (2004). Orientational manoeuvres in the dark: Dissociating allocentric and egocentric influences on spatial memory. *Cognition*, *94*(2), 149–166. https://doi.org/https://doi.org/10.1016/j.cognition.2004.01.001

Burkhardt, M., Bergelt, J., Gönner, L., Dinkelbach, H. Ü., Beuth, F., Schwarz, A., Bicanski, A., Burgess, N., & Hamker, F. H. (2023). A large-scale neuro-computational model of spatial cognition integrating memory with vision. *Neural Networks*, *167*, 473–488. https://doi.org/10.1016/j.neunet.2023.08.034

Buschke, H. (1984). Cued recall in Amnesia. *Journal of Clinical Neuropsychology*, *6*(4), 433–440. https://doi.org/10.1080/01688638408401233

Bush, D., Barry, C., & Burgess, N. (2014). What do grid cells contribute to place cell firing? *Trends in Neurosciences*, *37*(3), 136–145. https://doi.org/10.1016/j.tins.2013.12.003

Cahill, S., Pierce, M., Werner, P., Darley, A., & Bobersky, A. (2015). A Systematic Review of the Public's Knowledge and Understanding of Alzheimer's Disease and Dementia. *Alzheimer Disease & Associated Disorders*, *29*(3), 255–275. https://doi.org/10.1097/WAD.0000000000000102

Campbell, M. G., & Giocomo, L. M. (2018). Self-motion processing in visual and entorhinal cortices: Inputs, integration, and implications for position coding. *Journal of Neurophysiology*, *120*(4), 2091–2106. https://doi.org/10.1152/jn.00686.2017

Cardelli, L., Tullo, M. G., Galati, G., & Sulpizio, V. (2023). Effect of optic flow on spatial updating: Insight from an immersive virtual reality study. *Experimental Brain Research*, *241*(3), 865–874. https://doi.org/10.1007/s00221-023-06567-z

Cárdenas, A. M., Ardiles, A. O., Barraza, N., Baéz-Matus, X., & Caviedes, P. (2012). Role of Tau Protein in Neuronal Damage in Alzheimer's Disease and Down Syndrome. *Archives of Medical Research*, *43*(8), 645–654. https://doi.org/10.1016/j.arcmed.2012.10.012

Castegnaro, A. (2021, January). *Testing Spatial Cognition in Mild Cognitive Impairment Using Immersive Virtual Reality* [Doctoral]. UCL (University College London). Retrieved November 18, 2023, from https://discovery.ucl.ac.uk/id/eprint/10118890/

Castegnaro, A., Howett, D., Li, A., Harding, E., Chan, D., Burgess, N., & King, J. (2022). Assessing mild cognitive impairment using object-location memory in immersive virtual environments. *Hippocampus*, *32*(9), 660–678. https://doi.org/10.1002/hipo.23458

Castegnaro, A., Ji, Z., Rudzka, K., Chan, D., & Burgess, N. (2023). Overestimation in angular path integration precedes Alzheimer's dementia. *Current Biology*, *33*(21), 4650–4661.e7. https://doi.org/10.1016/j.cub.2023.09.047

Castillo Escamilla, J., León Estrada, I., Alcaraz-Iborra, M., & Cimadevilla Redondo, J. M. (2023). Aging: Working memory capacity and spatial strategies in a virtual orientation task. *GeroScience*, *45*(1), 159–175. https://doi.org/10.1007/s11357-022-00599-z

Castillo-Escamilla, J., Carmona, I., Salvador-Viñas, M. d. M., Frutos-Lorente, M., Ordoñez-Carrasco, J. L., & Cimadevilla, J. M. (2022). Clockwise rotation of perspective view improves spatial recognition of complex environments in aging. *Scientific Reports*, *12*(1), 18756. https://doi.org/10.1038/s41598-022-23301-x

Castro, D. M., Dillon, C., Machnicki, G., & Allegri, R. F. (2010). The economic cost of Alzheimer's disease: Family or public-health burden? *Dementia & Neuropsychologia*, *4*, 262–267. https://doi.org/10.1590/S1980-57642010DN40400003

Cavedo, E., Pievani, M., Boccardi, M., Galluzzi, S., Bocchetta, M., Bonetti, M., Thompson, P. M., & Frisoni, G. B. (2014). Medial temporal atrophy in early and late-onset Alzheimer's disease. *Neurobiology of Aging*, *35*(9), 2004–2012. https://doi.org/10.1016/j.neurobiolaging.2014.03.009

Cerami, C., Dubois, B., Boccardi, M., Monsch, A. U., Demonet, J. F., & Cappa, S. F. (2017). Clinical validity of delayed recall tests as a gateway biomarker for Alzheimer's disease in the context of a structured 5-phase development framework. *Neurobiology of Aging*, *52*, 153–166. https://doi.org/10.1016/j.neurobiolaging.2016.03.034

Chan, D., Gallaher, L. M., Moodley, K., Minati, L., Burgess, N., & Hartley, T. (2016). The 4 Mountains Test: A Short Test of Spatial Memory with High Sensitivity for the Diagnosis of Pre-dementia Alzheimer's Disease. *JoVE (Journal of Visualized Experiments)*, (116), e54454. https://doi.org/10.3791/54454

Chan, E., Baumann, O., Bellgrove, M., & Mattingley, J. (2012). From Objects to Landmarks: The Function of Visual Location Information in Spatial Navigation. *Frontiers in Psychology*, *3*. Retrieved November 26, 2023, from https://www.frontiersin.org/articles/10.3389/fpsyg.2012.00304

Chance, S. S., Gaunet, F., Beall, A. C., & Loomis, J. M. (1998). Locomotion Mode Affects the Updating of Objects Encountered During Travel: The Contribution of Vestibular and Proprioceptive Inputs to Path Integration. *Presence*, *7*(2), 168–178. https://doi.org/10.1162/105474698565659

Chau, S. A., Herrmann, N., Sherman, C., Chung, J., Eizenman, M., Kiss, A., & Lanctôt, K. L. (2017). Visual Selective Attention Toward Novel Stimuli Predicts Cognitive Decline in Alzheimer's Disease Patients. *Journal of Alzheimer's Disease*, *55*(4), 1339–1349. https://doi.org/10.3233/JAD-160641

Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). SMOTE: Synthetic Minority Over-sampling Technique. *Journal of Artificial Intelligence Research*, *16*, 321–357. https://doi.org/10.1613/jair.953

Chen, G., Lu, Y., King, J. A., Cacucci, F., & Burgess, N. (2019). Differential influences of environment and self-motion on place and grid cell firing. *Nature Communications*, *10*(1), 630. https://doi.org/10.1038/s41467-019-08550-1

Chen, X.-l., & Hou, W.-j. (2022, May). Identifying Fixation and Saccades in Virtual Reality. https://doi.org/10.48550/arXiv.2205.04121

Cherry, K. E., & Park, D. C. (1993). Individual difference and contextual variables influence spatial memory in younger and older adults. *Psychology and Aging*, *8*(4), 517–526. https://doi.org/10.1037/0882-7974.8.4.517

Chi, S. Y., Chua, E. F., Kieschnick, D. W., & Rabin, L. A. (2022). Retrospective metamemory monitoring of semantic memory in community-dwelling older adults with subjective cognitive decline and mild cognitive impairment. *Neuropsychological Rehabilitation*, *32*(3), 429–463. https://doi.org/10.1080/09602011.2020.1831552

Chua, E. F., Pergolizzi, D., & Weintraub, R. R. (2014). The Cognitive Neuroscience of Metamemory Monitoring: Understanding Metamemory Processes, Subjective Levels Expressed, and Metacognitive Accuracy. In S. M. Fleming & C. D. Frith (Eds.), *The Cognitive Neuroscience of Metacognition* (pp. 267–291). Springer. Retrieved November 26, 2023, from https://doi.org/10.1007/978-3-642-45190-4_12

Cipolotti, L., Bird, C., Good, T., Macmanus, D., Rudge, P., & Shallice, T. (2006). Recollection and familiarity in dense hippocampal amnesia: A case study. *Neuropsychologia*, *44*(3), 489–506. https://doi.org/10.1016/j.neuropsychologia.2005.05.014

Clay, V., König, P., & König, S. (2019). Eye Tracking in Virtual Reality. *Journal of Eye Movement Research*, *12*(1), 10.16910/jemr.12.1.3. https://doi.org/10.16910/jemr.12.1.3

Coco, M. I., Maruta, C., Pavão Martins, I., & Della Sala, S. (2023). Locations of objects are better remembered than their identities in naturalistic scenes: An eye-tracking experiment in mild cognitive impairment. *Neuropsychology*, *37*(7), 741–752. https://doi.org/10.1037/neu0000869

Coco, M. I., Merendino, G., Zappalà, G., & Della Sala, S. (2021). Semantic interference mechanisms on long-term visual memory and their eye-movement signatures in mild cognitive impairment. *Neuropsychology*, *35*(5), 498–513. https://doi.org/10.1037/neu0000734

Colombo, D., Serino, S., Tuena, C., Pedroli, E., Dakanalis, A., Cipresso, P., & Riva, G. (2017). Egocentric and allocentric spatial reference frames in aging: A systematic review. *Neuroscience & Biobehavioral Reviews*, *80*, 605–621. https://doi.org/10.1016/j.neubiorev.2017.07.012

Comblain, C., D'Argembeau, A., Van der Linden, M., & Aldenhoff, L. (2004). The effect of ageing on the recollection of emotional and neutral pictures [Publisher: Routledge _eprint: https://doi.org/10.1080/09658210344000477]. *Memory*, *12*(6), 673–684. https://doi.org/10.1080/09658210344000477

Cook, D., & Kesner, R. P. (1988). Caudate nucleus and memory for egocentric localization. *Behavioral and Neural Biology*, *49*(3), 332–343. https://doi.org/10.1016/S0163-1047(88)90338-X

Corwin, J., & Bylsma, F. W. (1993). Psychological examination of traumatic encephalopathy. *Clinical Neuropsychologist*, *7*(1), 3–21. https://doi.org/10.1080/13854049308401883

Cosentino, S., Metcalfe, J., Butterfield, B., & Stern, Y. (2007). Objective Metamemory Testing Captures Awareness of Deficit in Alzheimer's Disease. *Cortex*, *43*(7), 1004–1019. https://doi.org/10.1016/S0010-9452(08)70697-X

Coughlan, G., Laczó, J., Hort, J., Minihane, A.-M., & Hornberger, M. (2018). Spatial navigation deficits — overlooked cognitive marker for preclinical Alzheimer disease? *Nature Reviews Neurology*, *14*(8), 496–506. https://doi.org/10.1038/s41582-018-0031-x

Crutcher, M. D., Calhoun-Haney, R., Manzanares, C. M., Lah, J. J., Levey, A. I., & Zola, S. M. (2009). Eye Tracking During a Visual Paired Comparison Task as a Predictor of Early Dementia. *American Journal of Alzheimer's Disease & Other Dementias®*, *24*(3), 258–266. https://doi.org/10.1177/1533317509332093

Daffner, K. R., Scinto, L. F. M., Weintraub, S., Guinessey, J. E., & Mesulam, M. M. (1992). Diminished curiosity in patients with probable Alzheimer's disease as measured by exploratory eye movements. *Neurology*, *42*(2), 320–320. https://doi.org/10.1212/WNL.42.2.320

Das, V. E., Thomas, C. W., Zivotofsky, A. Z., & Leigh, R. J. (1996). Measuring Eye Movements During Locomotion: Filtering Techniques for Obtaining Velocity Signals From a Video-based Eye Monitor. *Journal of Vestibular Research*, *6*(6), 455–461. https://doi.org/10.3233/VES-1996-6606

De Beni, R., & Palladino, P. (2004). Decline in working memory updating through ageing: Intrusion error analyses. *Memory*, *12*(1), 75–89. https://doi.org/10.1080/09658210244000568

Deary, I. J., Corley, J., Gow, A. J., Harris, S. E., Houlihan, L. M., Marioni, R. E., Penke, L., Rafnsson, S. B., & Starr, J. M. (2009). Age-associated cognitive decline. *British Medical Bulletin*, *92*(1), 135–152. https://doi.org/10.1093/bmb/ldp033

Derby, C. A., Burns, L. C., Wang, C., Katz, M. J., Zimmerman, M. E., L'Italien, G., Guo, Z., Berman, R. M., & Lipton, R. B. (2013). Screening for predementia AD: Time-dependent operating characteristics of episodic memory tests. *Neurology*, *80*(14), 1307–1314. https://doi.org/10.1212/WNL.0b013e31828ab2c9

de Toledo-Morrell, L., Dickerson, B., Sullivan, M., Spanovic, C., Wilson, R., & Bennett, D. (2000). Hemispheric differences in hippocampal volume predict verbal and spatial memory performance in patients with Alzheimer's disease. *Hippocampus*, *10*(2), 136–142. https://doi.org/10.1002/(SICI)1098-1063(2000)10:2⟨136::AID-HIPO2⟩3.0.CO;2-J

de Toledo-Morrell, L., Morrell, F., & Fleming, S. (1984). Age-dependent deficits in spatial memory are related to impaired hippocampal kindling. *Behavioral Neuroscience*, *98*(5), 902–907. https://doi.org/10.1037/0735-7044.98.5.902

Dewhurst, R., Nyström, M., Jarodzka, H., Foulsham, T., Johansson, R., & Holmqvist, K. (2012). It depends on how you look at it: Scanpath comparison in multiple dimensions with MultiMatch, a vector-based approach. *Behavior Research Methods*, *44*(4), 1079–1100. https://doi.org/10.3758/s13428-012-0212-2

Diaz, G., Cooper, J., Kit, D., & Hayhoe, M. (2013). Real-time recording and classification of eye movements in an immersive virtual environment. *Journal of Vision*, *13*(12), 5. https://doi.org/10.1167/13.12.5

Dierckx, E., Engelborghs, S., Raedt, R. D., Deyn, P. P. D., & Ponjaert-Kristoffersen, I. (2007). Differentiation between mild cognitive impairment, Alzheimer's disease and depression by means of cued recall. *Psychological Medicine*, *37*(5), 747–755. https://doi.org/10.1017/S003329170600955X

Dodson, C. S., Bawa, S., & Krueger, L. E. (2007). Aging, metamemory, and high-confidence errors: A misrecollection account. *Psychology and Aging*, *22*(1), 122–133. https://doi.org/10.1037/0882-7974.22.1.122

Dodson, C. S., Spaniol, M., O'Connor, M. K., Deason, R. G., Ally, B. A., & Budson, A. E. (2011). Alzheimer's disease and memory-monitoring impairment: Alzheimer's patients show a monitoring deficit that is greater

than their accuracy deficit. *Neuropsychologia*, *49*(9), 2609–2618. https://doi.org/10.1016/j.neuropsychologia.2011.05.008

Doeller, C. F., Barry, C., & Burgess, N. (2010). Evidence for grid cells in a human memory network. *Nature*, *463*(7281), 657–661. https://doi.org/10.1038/nature08704

Doeller, C. F., King, J. A., & Burgess, N. (2008). Parallel striatal and hippocampal systems for landmarks and boundaries in spatial memory. *Proceedings of the National Academy of Sciences*, *105*(15), 5915–5920. https://doi.org/10.1073/pnas.0801489105

Downs, S. H., & Black, N. (1998). The feasibility of creating a checklist for the assessment of the methodological quality both of randomised and non-randomised studies of health care interventions. *Journal of Epidemiology & Community Health*, *52*(6), 377–384. https://doi.org/10.1136/jech.52.6.377

Drewes, J., Feder, S., & Einhäuser, W. (2021). Gaze During Locomotion in Virtual Reality and the Real World. *Frontiers in Neuroscience*, *15*. Retrieved November 26, 2023, from https://www.frontiersin.org/articles/10.3389/fnins.2021.656913

Dubois, B., Villain, N., Frisoni, G. B., Rabinovici, G. D., Sabbagh, M., Cappa, S., Bejanin, A., Bombois, S., Epelbaum, S., Teichmann, M., Habert, M.-O., Nordberg, A., Blennow, K., Galasko, D., Stern, Y., Rowe, C. C., Salloway, S., Schneider, L. S., Cummings, J. L., & Feldman, H. H. (2021). Clinical diagnosis of Alzheimer's disease: Recommendations of the International Working Group. *The Lancet Neurology*, *20*(6), 484–496. https://doi.org/10.1016/S1474-4422(21)00066-1

Ebeid, I. A., & Gwizdka, J. (2018). Real-time gaze transition entropy. *Proceedings of the 2018 ACM Symposium on Eye Tracking Research & Applications*, 1–3. https://doi.org/10.1145/3204493.3208340

Einhäuser, W., Kruse, W., Hoffmann, K.-P., & König, P. (2006). Differences of monkey and human overt attention under natural conditions. *Vision Research*, *46*(8), 1194–1209. https://doi.org/10.1016/j.visres.2005.08.032

Ekstrom, A. D., Arnold, A. E. G. F., & Iaria, G. (2014). A critical review of the allocentric spatial representation and its neural underpinnings: Toward a network-based perspective. *Frontiers in Human Neuroscience*, *8*. Retrieved November 25, 2023, from https://www.frontiersin.org/articles/10.3389/fnhum.2014.00803

Ekstrom, A. D., Kahana, M. J., Caplan, J. B., Fields, T. A., Isham, E. A., Newman, E. L., & Fried, I. (2003). Cellular networks underlying human spatial navigation. *Nature*, *425*(6954), 184–188. https://doi.org/10.1038/nature01964

Ellmore, T. M., & McNaughton, B. L. (2004). Human Path Integration by Optic Flow. *Spatial Cognition & Computation*, *4*(3), 255–272. https://doi.org/10.1207/s15427633scc0403_3

Emam, S., Nasrollahpour, M., Colarusso, B., Cai, X., Grant, S., Kulkarni, P., Ekenseair, A., Gharagouzloo, C., Ferris, C. F., & Sun, N.-X. (2020). Detection of presymptomatic Alzheimer's disease through breath biomarkers. *Alzheimer's & Dementia: Diagnosis, Assessment & Disease Monitoring*, *12*(1), e12088. https://doi.org/10.1002/dad2.12088

Enders, L. R., Smith, R. J., Gordon, S. M., Ries, A. J., & Touryan, J. (2021). Gaze Behavior During Navigation and Visual Search of an Open-World Virtual Environment. *Frontiers in Psychology*, *12*. Retrieved November 26, 2023, from https://www.frontiersin.org/articles/10.3389/fpsyg.2021.681042

Epstein, R. A., Patai, E. Z., Julian, J. B., & Spiers, H. J. (2017). The cognitive map in humans: Spatial navigation and beyond. *Nature Neuroscience*, *20*(11), 1504–1513. https://doi.org/10.1038/nn.4656

Estévez-González, A., Kulisevsky, J., Boltes, A., Otermín, P., & García-Sánchez, C. (2003). Rey verbal learning test is a useful tool for differential diagno-

sis in the preclinical phase of Alzheimer's disease: Comparison with mild cognitive impairment and normal aging. *International Journal of Geriatric Psychiatry*, *18*(11), 1021–1028. https://doi.org/10.1002/gps.1010

Evans, T., Bicanski, A., Bush, D., & Burgess, N. (2016). How environment and self-motion combine in neural representations of space. *The Journal of Physiology*, *594*(22), 6535–6546. https://doi.org/10.1113/JP270666

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*(2), 175–191. https://doi.org/10.3758/BF03193146

Fehlmann, B., Coynel, D., Schicktanz, N., Milnik, A., Gschwind, L., Hofmann, P., Papassotiropoulos, A., & de Quervain, D. J.-F. (2020). Visual Exploration at Higher Fixation Frequency Increases Subsequent Memory Recall. *Cerebral Cortex Communications*, *1*(1), tgaa032. https://doi.org/10.1093/texcom/tgaa032

Fernández, G., Castro, L. R., Schumacher, M., & Agamennoni, O. E. (2015). Diagnosis of mild Alzheimer disease through the analysis of eye movements during reading. *Journal of Integrative Neuroscience*, *14*(01), 121–133. https://doi.org/10.1142/S0219635215500090

Fischer, P., Jungwirth, S., Zehetmayer, S., Weissgram, S., Hoenigschnabl, S., Gelpi, E., Krampla, W., & Tragl, K. H. (2007). Conversion from subtypes of mild cognitive impairment to Alzheimer dementia. *Neurology*, *68*(4), 288–291. https://doi.org/10.1212/01.wnl.0000252358.03285.9d

Flicker, C., Ferris, S. H., & Reisberg, B. (1993). A Two-Year Longitudinal Study of Cognitive Function in Normal Aging and Alzheimer's Disease. *Journal of Geriatric Psychiatry and Neurology*, *6*(2), 84–96. https://doi.org/10.1177/089198879300600205

Folstein, M. F., Folstein, S. E., & McHugh, P. R. (1975). "Mini-mental state": A practical method for grading the cognitive state of patients for the clini-

cian. *Journal of Psychiatric Research*, *12*(3), 189–198. https://doi.org/10.1016/0022-3956(75)90026-6

Ford, A., White, C. T., & Lichtenstein, M. (1959). Analysis of Eye Movements during Free Search. *JOSA*, *49*(3), 287–292. https://doi.org/10.1364/JOSA.49.000287

Fortenbaugh, F. C., Hicks, J. C., Hao, L., & Turano, K. A. (2007). Losing sight of the bigger picture: Peripheral field loss compresses representations of space. *Vision Research*, *47*(19), 2506–2520. https://doi.org/10.1016/j.visres.2007.06.012

Fortenbaugh, F. C., Hicks, J. C., & Turano, K. A. (2008). The Effect of Peripheral Visual Field Loss on Representations of Space: Evidence for Distortion and Adaptation. *Investigative Ophthalmology & Visual Science*, *49*(6), 2765–2772. https://doi.org/10.1167/iovs.07-1021

Foulsham, T., & Kingstone, A. (2013). Fixation-dependent memory for natural scenes: An experimental test of scanpath theory. *Journal of Experimental Psychology: General*, *142*(1), 41–56. https://doi.org/10.1037/a0028227

Frey, B. J., & Dueck, D. (2007). Clustering by Passing Messages Between Data Points. *Science*, *315*(5814), 972–976. https://doi.org/10.1126/science.1136800

Galeone, F., Pappalardo, S., Chieffi, S., Iavarone, A., & Carlomagno, S. (2011). Anosognosia for memory deficit in amnestic mild cognitive impairment and Alzheimer's disease. *International Journal of Geriatric Psychiatry*, *26*(7), 695–701. https://doi.org/10.1002/gps.2583

Gallo, D. A., Cramer, S. J., Wong, J. T., & Bennett, D. A. (2012). Alzheimer's disease can spare local metacognition despite global anosognosia: Revisiting the confidence–accuracy relationship in episodic memory. *Neuropsychologia*, *50*(9), 2356–2364. https://doi.org/10.1016/j.neuropsychologia.2012.06.005

Gao, F., Dai, L., Wang, Q., Liu, C., Deng, K., Cheng, Z., Lv, X., Wu, Y., Zhang, Z., Tao, Q., Yuan, J., Li, S., Wang, Y., Su, Y., Cheng, X., Ni, J., Wu, Z., Zhang, S., Shi, J., … China Aging and Neurodegenerative Initiative (CANDI) Consortium. (2023). Blood-based biomarkers for Alzheimer's disease: A multicenter-based cross-sectional and longitudinal study in China. *Science Bulletin*, *68*(16), 1800–1808. https://doi.org/10.1016/j.scib.2023.07.009

Garre-Olmo, J. (2018). Epidemiology of Alzheimer's disease and other dementias. *Revista de neurologia*, *66*(11), 377–386.

Geslani, D. M., Tierney, M. C., Herrmann, N., & Szalai, J. P. (2005). Mild Cognitive Impairment: An Operational Definition and Its Conversion Rate to Alzheimer's Disease. *Dementia and Geriatric Cognitive Disorders*, *19*(5-6), 383–389. https://doi.org/10.1159/000084709

Ghose, U., Srinivasan, A. A., Boyce, W. P., Xu, H., & Chng, E. S. (2020). PyTrack: An end-to-end analysis toolkit for eye tracking. *Behavior Research Methods*, *52*(6), 2588–2603. https://doi.org/10.3758/s13428-020-01392-6

Grande, G., Vanacore, N., Vetrano, D. L., Cova, I., Rizzuto, D., Mayer, F., Maggiore, L., Ghiretti, R., Cucumo, V., Mariani, C., Cappa, S. F., & Pomati, S. (2018). Free and cued selective reminding test predicts progression to Alzheimer's disease in people with mild cognitive impairment. *Neurological Sciences*, *39*(11), 1867–1875. https://doi.org/10.1007/s10072-018-3507-y

Gras, D., Daniel, M.-P., Labiale, G., Piolino, P., & Gyselinck, V. (2012). Effet du vieillissement sur la mémorisation d'itinéraires réels : L'importance de la mémoire de travail et de la mémoire épisodique. *Gériatrie et Psychologie Neuropsychiatrie du Vieillissement*, *10*(4), 463–470. https://doi.org/10.1684/pnv.2012.0370

Greenaway, M. C., Lacritz, L. H., Binegar, D., Weiner, M. F., Lipton, A., & Munro Cullum, C. (2006). Patterns of Verbal Memory Performance in Mild Cog-

nitive Impairment, Alzheimer Disease, and Normal Aging. *Cognitive and Behavioral Neurology*, *19*(2), 79. https://doi.org/10.1097/01.wnn.0000208290.57370.a3

Grossi, D., Becker, J. T., Smith, C., & Trojano, L. (1993). Memory for visuospatial patterns in Alzheimer's disease. *Psychological Medicine*, *23*(1), 65–70. https://doi.org/10.1017/S003329170003885X

Guarch, J., Marcos, T., Salamero, M., Gastó, C., & Blesa, R. (2008). Mild cognitive impairment: A risk indicator of later dementia, or a preclinical phase of the disease? *International Journal of Geriatric Psychiatry*, *23*(3), 257–265. https://doi.org/10.1002/gps.1871

Guest, F. L., Rahmoune, H., & Guest, P. C. (2020). Early Diagnosis and Targeted Treatment Strategy for Improved Therapeutic Outcomes in Alzheimer's Disease. In P. C. Guest (Ed.), *Reviews on New Drug Targets in Age-Related Disorders* (pp. 175–191). Springer International Publishing. Retrieved November 24, 2023, from https://doi.org/10.1007/978-3-030-42667-5_8

Gunawardena, N., Ginige, J. A., & Javadi, B. (2022). Eye-tracking Technologies in Mobile Devices Using Edge Computing: A Systematic Review. *ACM Computing Surveys*, *55*(8), 158:1–158:33. https://doi.org/10.1145/3546938

Hafting, T., Fyhn, M., Molden, S., Moser, M.-B., & Moser, E. I. (2005). Microstructure of a spatial map in the entorhinal cortex. *Nature*, *436*(7052), 801–806. https://doi.org/10.1038/nature03721

Hannula, D. E., Ryan, J. D., Tranel, D., & Cohen, N. J. (2007). Rapid Onset Relational Memory Effects Are Evident in Eye Movement Behavior, but Not in Hippocampal Amnesia. *Journal of Cognitive Neuroscience*, *19*(10), 1690–1705. https://doi.org/10.1162/jocn.2007.19.10.1690

Hardy, J., & Selkoe, D. J. (2002). The Amyloid Hypothesis of Alzheimer's Disease: Progress and Problems on the Road to Therapeutics. *Science*, *297*(5580), 353–356. https://doi.org/10.1126/science.1072994

Harris, M., Wiener, J., & Wolbers, T. (2012). Aging specifically impairs switching to an allocentric navigational strategy. *Frontiers in Aging Neuroscience*, *4*. Retrieved November 22, 2023, from https://www.frontiersin.org/articles/10.3389/fnagi.2012.00029

Hartley, T., Bird, C. M., Chan, D., Cipolotti, L., Husain, M., Vargha-Khadem, F., & Burgess, N. (2007). The hippocampus is required for short-term topographical memory in humans. *Hippocampus*, *17*(1), 34–48. https://doi.org/10.1002/hipo.20240

Hartley, T., Lever, C., Burgess, N., & O'Keefe, J. (2014). Space in the brain: How the hippocampal formation supports spatial cognition. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *369*(1635), 20120510. https://doi.org/10.1098/rstb.2012.0510

Hayes, T. R., Petrov, A. A., & Sederberg, P. B. (2011). A novel method for analyzing sequential eye movements reveals strategic influence on Raven's Advanced Progressive Matrices. *Journal of Vision*, *11*(10), 10. https://doi.org/10.1167/11.10.10

Hess, T. M., & Ennis, G. E. (2012). Age Differences in the Effort and Costs Associated With Cognitive Activity. *The Journals of Gerontology: Series B*, *67*(4), 447–455. https://doi.org/10.1093/geronb/gbr129

Heywood-Everett, E., Baker, D. H., & Hartley, T. (2022). Testing the precision of spatial memory representations using a change-detection task: Effects of viewpoint change. *Journal of Cognitive Psychology*, *34*(1), 127–141. https://doi.org/10.1080/20445911.2020.1863414

Hicks, S. L., Sharma, R., Khan, A. N., Berna, C. M., Waldecker, A., Talbot, K., Kennard, C., & Turner, M. R. (2013). An Eye-Tracking Version of the Trail-

Making Test. *PLOS ONE*, *8*(12), e84061. https://doi.org/10.1371/journal.pone.0084061

Hilton, C., Muffato, V., Slattery, T. J., Miellet, S., & Wiener, J. (2020). Differences in Encoding Strategy as a Potential Explanation for Age-Related Decline in Place Recognition Ability. *Frontiers in Psychology*, *11*. Retrieved September 29, 2023, from https://www.frontiersin.org/articles/10.3389/fpsyg.2020.02182

Hirose, Y., Kennedy, A., & Tatler, B. W. (2010). Perception and memory across viewpoint changes in moving images. *Journal of Vision*, *10*(4), 2. https://doi.org/10.1167/10.4.2

Holm, L., & Mäntylä, T. (2007). Memory for scenes: Refixations reflect retrieval. *Memory & Cognition*, *35*(7), 1664–1674. https://doi.org/10.3758/BF03193500

Holm, S. (1979). A Simple Sequentially Rejective Multiple Test Procedure. *Scandinavian Journal of Statistics*, *6*(2), 65–70. Retrieved November 22, 2023, from https://www.jstor.org/stable/4615733

Holmes, C. A., Newcombe, N. S., & Shipley, T. F. (2018). Move to learn: Integrating spatial information from multiple viewpoints. *Cognition*, *178*, 7–25. https://doi.org/10.1016/j.cognition.2018.05.003

Holmqvist, K., & Blignaut, P. (2020). Small eye movements cannot be reliably measured by video-based P-CR eye-trackers. *Behavior Research Methods*, *52*(5), 2098–2121. https://doi.org/10.3758/s13428-020-01363-x

Hooge, I. T. C., Niehorster, D. C., Nyström, M., Andersson, R., & Hessels, R. S. (2018). Is human classification by experienced untrained observers a gold standard in fixation detection? *Behavior Research Methods*, *50*(5), 1864–1881. https://doi.org/10.3758/s13428-017-0955-x

Hoppe, D., & Rothkopf, C. A. (2019). Multi-step planning of eye movements in visual search. *Scientific Reports*, *9*(1), 144. https://doi.org/10.1038/s41598-018-37536-0

Hort, J., Laczó, J., Vyhnálek, M., Bojar, M., Bureš, J., & Vlček, K. (2007). Spatial navigation deficit in amnestic mild cognitive impairment. *Proceedings of the National Academy of Sciences*, *104*(10), 4042–4047. https://doi.org/10.1073/pnas.0611314104

Howett, D., Castegnaro, A., Krzywicka, K., Hagman, J., Marchment, D., Henson, R., Rio, M., King, J. A., Burgess, N., & Chan, D. (2019). Differentiation of mild cognitive impairment using an entorhinal cortex-based test of virtual reality navigation. *Brain*, *142*(6), 1751–1766. https://doi.org/https://doi.org/10.1093/brain/awz116

HTC. (n.d.). VIVE Pro Eye Overview — VIVE Southeast Asia. Retrieved October 1, 2023, from https://www.vive.com/sea/product/vive-pro-eye/overview/

Hubert, L., & Arabie, P. (1985). Comparing partitions. *Journal of Classification*, *2*(1), 193–218. https://doi.org/10.1007/BF01908075

Huff, M., Papenmeier, F., Jahn, G., & Hesse, F. W. (2010). Eye movements across viewpoint changes in multiple object tracking. *Visual Cognition*, *18*(9), 1368–1391. https://doi.org/10.1080/13506285.2010.495878

Iachini, T., & Ruggiero, G. (2006). Egocentric and allocentric spatial frames of reference: A direct measure. *Cognitive Processing*, *7*(1), 126–127. https://doi.org/10.1007/s10339-006-0100-8

Isler, R. B., Parsonson, B. S., & Hansson, G. J. (1997). Age related effects of restricted head movements on the useful field of view of drivers. *Accident Analysis & Prevention*, *29*(6), 793–801. https://doi.org/10.1016/S0001-4575(97)00048-1

Jack, C. R. (2022). Advances in Alzheimer's disease research over the past two decades. *The Lancet Neurology*, *21*(10), 866–869. https://doi.org/10.1016/S1474-4422(22)00298-8

Jack, C. R., Bennett, D. A., Blennow, K., Carrillo, M. C., Dunn, B., Haeberlein, S. B., Holtzman, D. M., Jagust, W., Jessen, F., Karlawish, J., Liu, E., Molinuevo, J. L., Montine, T., Phelps, C., Rankin, K. P., Rowe, C. C.,

Scheltens, P., Siemers, E., Snyder, H. M., . . . Silverberg, N. (2018). NIA-AA Research Framework: Toward a biological definition of Alzheimer's disease. *Alzheimer's & Dementia*, *14*(4), 535–562. https://doi.org/10.1016/j.jalz.2018.02.018

Jack, C. R., Knopman, D. S., Jagust, W. J., Shaw, L. M., Aisen, P. S., Weiner, M. W., Petersen, R. C., & Trojanowski, J. Q. (2010). Hypothetical model of dynamic biomarkers of the Alzheimer's pathological cascade. *The Lancet Neurology*, *9*(1), 119–128. https://doi.org/10.1016/S1474-4422(09)70299-6

Jack, C. R., Petersen, R. C., Xu, Y., O'Brien, P. C., Smith, G. E., Ivnik, R. J., Tangalos, E. G., & Kokmen, E. (1998). Rate of medial temporal lobe atrophy in typical aging and Alzheimer's disease. *Neurology*, *51*(4), 993–999. https://doi.org/10.1212/WNL.51.4.993

Jack, C. R., Wiste, H. J., Therneau, T. M., Weigand, S. D., Knopman, D. S., Mielke, M. M., Lowe, V. J., Vemuri, P., Machulda, M. M., Schwarz, C. G., Gunter, J. L., Senjem, M. L., Graff-Radford, J., Jones, D. T., Roberts, R. O., Rocca, W. A., & Petersen, R. C. (2019). Associations of Amyloid, Tau, and Neurodegeneration Biomarker Profiles With Rates of Memory Decline Among Individuals Without Dementia. *JAMA*, *321*(23), 2316–2325. https://doi.org/10.1001/jama.2019.7437

Jacobs, J., Kahana, M. J., Ekstrom, A. D., Mollison, M. V., & Fried, I. (2010). A sense of direction in human entorhinal cortex. *Proceedings of the National Academy of Sciences*, *107*(14), 6487–6492. https://doi.org/10.1073/pnas.0911213107

Jacobs, J., Weidemann, C. T., Miller, J. F., Solway, A., Burke, J. F., Wei, X.-X., Suthana, N., Sperling, M. R., Sharan, A. D., Fried, I., & Kahana, M. J. (2013). Direct recordings of grid-like neuronal activity in human spatial navigation. *Nature Neuroscience*, *16*(9), 1188–1190. https://doi.org/10.1038/nn.3466

Jarodzka, H., Holmqvist, K., & Nyström, M. (2010). A vector-based, multidimensional scanpath similarity measure. *Proceedings of the 2010 Symposium on Eye-Tracking Research & Applications*, 211–218. https://doi.org/10.1145/1743666.1743718

Jessen, F., Amariglio, R. E., Buckley, R. F., Flier, W. M. v. d., Han, Y., Molinuevo, J. L., Rabin, L., Rentz, D. M., Rodriguez-Gomez, O., Saykin, A. J., Sikkes, S. A. M., Smart, C. M., Wolfsgruber, S., & Wagner, M. (2020). The characterisation of subjective cognitive decline. *The Lancet Neurology*, *19*(3), 271–278. https://doi.org/10.1016/S1474-4422(19)30368-0

Johnson, M. K., Kuhl, B. A., Mitchell, K. J., Ankudowich, E., & Durbin, K. A. (2015). Age-related differences in the neural basis of the subjective vividness of memories: Evidence from multivoxel pattern classification. *Cognitive, Affective, & Behavioral Neuroscience*, *15*(3), 644–661. https://doi.org/10.3758/s13415-015-0352-9

Kapoula, Z., Yang, Q., Otero-Millan, J., Xiao, S., Macknik, S. L., Lang, A., Verny, M., & Martinez-Conde, S. (2014). Distinctive features of microsaccades in Alzheimer's disease and in mild cognitive impairment. *AGE*, *36*(2), 535–543. https://doi.org/10.1007/s11357-013-9582-3

Kavcic, V., Fernandez, R., Logan, D., & Duffy, C. J. (2006). Neurophysiological and perceptual correlates of navigational impairment in Alzheimer's disease. *Brain*, *129*(3), 736–746. https://doi.org/10.1093/brain/awh727

Kehoe, E. G., McNulty, J. P., Mullins, P. G., & Bokde, A. L. W. (2014). Advances in MRI biomarkers for the diagnosis of Alzheimer's disease. *Biomarkers in Medicine*, *8*(9), 1151–1169. https://doi.org/10.2217/bmm.14.42

Khan, U. A., Liu, L., Provenzano, F. A., Berman, D. E., Profaci, C. P., Sloan, R., Mayeux, R., Duff, K. E., & Small, S. A. (2014). Molecular drivers and cortical spread of lateral entorhinal cortex dysfunction in preclinical Alzheimer's disease. *Nature Neuroscience*, *17*(2), 304–311. https://doi.org/10.1038/nn.3606

Kheirkhah, A., Azadeh, A., Saberi, M., Azaron, A., & Shakouri, H. (2013). Improved estimation of electricity demand function by using of artificial neural network, principal component analysis and data envelopment analysis. *Computers & Industrial Engineering*, *64*(1), 425–441. https://doi.org/10.1016/j.cie.2012.09.017

Killian, N. J., Jutras, M. J., & Buffalo, E. A. (2012). A map of visual space in the primate entorhinal cortex. *Nature*, *491*(7426), 761–764. https://doi.org/10.1038/nature11587

Kim, K. W., Choi, J., Chin, J., Lee, B. H., & Na, D. L. (2022). Eye-Tracking Metrics for Figure-Copying Processes in Early- vs. Late-Onset Alzheimer's Disease. *Frontiers in Neurology*, *13*. Retrieved September 28, 2023, from https://www.frontiersin.org/articles/10.3389/fneur.2022.844341

Kim, M., & Maguire, E. A. (2019). Can we study 3D grid codes non-invasively in the human brain? Methodological considerations and fMRI findings. *NeuroImage*, *186*, 667–678. https://doi.org/10.1016/j.neuroimage.2018.11.041

King, J. A., Burgess, N., Hartley, T., Vargha-Khadem, F., & O'Keefe, J. (2002). Human hippocampus and viewpoint dependence in spatial memory. *Hippocampus*, *12*(6), 811–820. https://doi.org/10.1002/hipo.10070

Kirova, A.-M., Bays, R. B., & Lagalwar, S. (2015). Working Memory and Executive Function Decline across Normal Aging, Mild Cognitive Impairment, and Alzheimer's Disease. *BioMed Research International*, *2015*, e748212. https://doi.org/10.1155/2015/748212

Klatzky, R. L. (1998). Allocentric and Egocentric Spatial Representations: Definitions, Distinctions, and Interconnections. In C. Freksa, C. Habel, & K. F. Wender (Eds.), *Spatial Cognition: An Interdisciplinary Approach to Representing and Processing Spatial Knowledge* (pp. 1–17). Springer. Retrieved November 25, 2023, from https://doi.org/10.1007/3-540-69342-4_1

Knight, M. J., Wearn, A., Coulthard, E., & Kauppinen, R. A. (2019). T2 Relaxometry and Diffusion Tensor Indices of the Hippocampus and Entorhinal Cortex Improve Sensitivity and Specificity of MRI to Detect Amnestic Mild Cognitive Impairment and Alzheimer's Disease Dementia. *Journal of Magnetic Resonance Imaging*, *49*(2), 445–455. https://doi.org/10.1002/jmri.26195

Krejtz, K., Duchowski, A., Szmidt, T., Krejtz, I., González Perilli, F., Pires, A., Vilaro, A., & Villalobos, N. (2015). Gaze Transition Entropy. *ACM Transactions on Applied Perception*, *13*(1), 4:1–4:20. https://doi.org/10.1145/2834121

Krejtz, K., Szmidt, T., Duchowski, A. T., & Krejtz, I. (2014). Entropy-based statistical analysis of eye movement transitions. *Proceedings of the Symposium on Eye Tracking Research and Applications*, 159–166. https://doi.org/10.1145/2578153.2578176

Kropff, E., Carmichael, J. E., Moser, M.-B., & Moser, E. I. (2015). Speed cells in the medial entorhinal cortex. *Nature*, *523*(7561), 419–424. https://doi.org/10.1038/nature14622

Kumar, A., Timmermans, N., Burch, M., & Mueller, K. (2019). Clustered eye movement similarity matrices. *Proceedings of the 11th ACM Symposium on Eye Tracking Research & Applications*, 1–9. https://doi.org/10.1145/3317958.3319811

Laczó, J., Andel, R., Vyhnalek, M., Vlcek, K., Nedelska, Z., Matoska, V., Gazova, I., Mokrisova, I., Sheardova, K., & Hort, J. (2014). APOE and spatial navigation in amnestic MCI: Results from a computer-based test. *Neuropsychology*, *28*(5), 676–684. https://doi.org/10.1037/neu0000072

Lafleche, G., & Albert, M. S. (1995). Executive function deficits in mild Alzheimer's disease. *Neuropsychology*, *9*(3), 313–320. https://doi.org/10.1037/0894-4105.9.3.313

Lagun, D., Manzanares, C., Zola, S. M., Buffalo, E. A., & Agichtein, E. (2011). Detecting cognitive impairment by eye movement analysis using automatic classification algorithms. *Journal of Neuroscience Methods*, *201*(1), 196–203. https://doi.org/10.1016/j.jneumeth.2011.06.027

Lam, A. T. A. D., Axmacher, N., Fell, J., Staresina, B. P., Gauggel, S., Wagner, T., Olligs, J., & Weis, S. (2012). Monitoring the Mind: The Neurocognitive Correlates of Metamemory. *PLOS ONE*, *7*(1), e30009. https://doi.org/10.1371/journal.pone.0030009

Laske, C. (2014). Phase 3 Trials of Solanezumab and Bapineuzumab for Alzheimer's Disease. *New England Journal of Medicine*, *370*(15), 1459–1460. https://doi.org/10.1056/NEJMc1402193

Laurens, B., Planche, V., Cubizolle, S., Declerck, L., Dupouy, S., Formaglio, M., Koric, L., Seassau, M., Tilikete, C., Vighetto, A., Ceccaldi, M., & Tison, F. (2019). A Spatial Decision Eye-Tracking Task in Patients with Prodromal and Mild Alzheimer's Disease. *Journal of Alzheimer's Disease*, *71*(2), 613–621. https://doi.org/10.3233/JAD-190549

Lee, J.-Y., Kho, S., Yoo, H. B., Park, S., Choi, J.-S., Kwon, J. S., Cha, K. R., & Jung, H.-Y. (2014). Spatial memory impairments in amnestic mild cognitive impairment in a virtual radial arm maze. *Neuropsychiatric Disease and Treatment*, 653. https://doi.org/10.2147/NDT.S58185

Lee, Y., Jung, K.-T., & Lee, H.-C. (2022). Use of gaze entropy to evaluate situation awareness in emergency accident situations of nuclear power plant. *Nuclear Engineering and Technology*, *54*(4), 1261–1270. https://doi.org/10.1016/j.net.2021.10.022

Lemaıtre, G., & Nogueira, F. (2017). Imbalanced-learn: A Python Toolbox to Tackle the Curse of Imbalanced Datasets in Machine Learning. *The Journal of Machine Learning Research*, *18*(1), 559–563.

Lester, A. W., Moffat, S. D., Wiener, J. M., Barnes, C. A., & Wolbers, T. (2017). The Aging Navigational System. *Neuron*, *95*(5), 1019–1035. https://doi.org/10.1016/j.neuron.2017.06.037

Li, Q., Pan, F.-F., Huang, Q., Lo, C.-Y. Z., Xie, F., & Guo, Q. (2022). Altered metamemory precedes cognitive impairment in subjective cognitive decline with positive amyloid-beta. *Frontiers in Aging Neuroscience*, *14*. Retrieved October 7, 2023, from https://www.frontiersin.org/articles/10.3389/fnagi.2022.1046445

Li, Q., Sun, X., Cui, L., Zhong, Y., Wang, B., Miao, Y., Hu, X., & Guo, Q. (2022). Alterations in metamemory capacity and neural correlates in a subtype of subjective cognitive decline. *NeuroImage: Clinical*, *36*, 103255. https://doi.org/10.1016/j.nicl.2022.103255

Liang, Y., Pertzov, Y., Nicholas, J. M., Henley, S. M. D., Crutch, S., Woodward, F., Leung, K., Fox, N. C., & Husain, M. (2016). Visual short-term memory binding deficit in familial Alzheimer's disease. *Cortex*, *78*, 150–164. https://doi.org/10.1016/j.cortex.2016.01.015

Lipinska, B., & Bäckman, L. (1996). Feeling-of-knowing in fact retrieval: Further evidence for preservation in early Alzheimer's disease. *Journal of the International Neuropsychological Society*, *2*(4), 350–358. https://doi.org/10.1017/S1355617700001375

Liu, Z.-X., Shen, K., Olsen, R. K., & Ryan, J. D. (2017). Visual Sampling Predicts Hippocampal Activity. *Journal of Neuroscience*, *37*(3), 599–609. https://doi.org/10.1523/JNEUROSCI.2610-16.2016

Livingston, G., Huntley, J., Sommerlad, A., Ames, D., Ballard, C., Banerjee, S., Brayne, C., Burns, A., Cohen-Mansfield, J., Cooper, C., Costafreda, S. G., Dias, A., Fox, N., Gitlin, L. N., Howard, R., Kales, H. C., Kivimäki, M., Larson, E. B., Ogunniyi, A., ... Mukadam, N. (2020). Dementia prevention, intervention, and care: 2020 report of the Lancet Commission.

The Lancet, *396*(10248), 413–446. https://doi.org/10.1016/S0140-6736(20)30367-6

Llanes-Jurado, J., Marín-Morales, J., Guixeres, J., & Alcañiz, M. (2020). Development and Calibration of an Eye-Tracking Fixation Identification Algorithm for Immersive Virtual Reality. *Sensors*, *20*(17), 4956. https://doi.org/10.3390/s20174956

Lloyd, S. (1982). Least squares quantization in PCM. *IEEE Transactions on Information Theory*, *28*(2), 129–137. https://doi.org/10.1109/TIT.1982.1056489

Loftus, G. R. (1972). Eye fixations and recognition memory for pictures. *Cognitive Psychology*, *3*(4), 525–551. https://doi.org/10.1016/0010-0285(72)90021-7

Lohr, D. J., Friedman, L., & Komogortsev, O. V. (2019, December). Evaluating the Data Quality of Eye Tracking Signals from a Virtual Reality System: Case Study using SMI's Eye-Tracking HTC Vive. https://doi.org/10.48550/arXiv.1912.02083

Lopera, F., Marino, C., Chandrahas, A. S., O'Hare, M., Villalba-Moreno, N. D., Aguillon, D., Baena, A., Sanchez, J. S., Vila-Castelar, C., Ramirez Gomez, L., Chmielewska, N., Oliveira, G. M., Littau, J. L., Hartmann, K., Park, K., Krasemann, S., Glatzel, M., Schoemaker, D., Gonzalez-Buendia, L., … Quiroz, Y. T. (2023). Resilience to autosomal dominant Alzheimer's disease in a Reelin-COLBOS heterozygous man. *Nature Medicine*, *29*(5), 1243–1252. https://doi.org/10.1038/s41591-023-02318-3

López, O. L., & DeKosky, S. T. (2008, January). Clinical symptoms in Alzheimer's disease. In *Handbook of Clinical Neurology* (pp. 207–216, Vol. 89). Elsevier. Retrieved November 24, 2023, from https://www.sciencedirect.com/science/article/pii/S0072975207012195

Lowin, A., Knapp, M., & McCrone, P. (2001). Alzheimer's disease in the UK: Comparative evidence on cost of illness and volume of health services research funding. *International Journal of Geriatric Psychiatry*, *16*(12), 1143–1148. https://doi.org/10.1002/gps.499

Lucas, H. D., Duff, M. C., & Cohen, N. J. (2018). The Hippocampus Promotes Effective Saccadic Information Gathering in Humans. *Journal of Cognitive Neuroscience*, *31*(2), 186–201. https://doi.org/10.1162/jocn_a_01336

Maass, A., Berron, D., Libby, L. A., Ranganath, C., & Düzel, E. (2015). Functional subregions of the human entorhinal cortex (H. Eichenbaum, Ed.). *eLife*, *4*, e06426. https://doi.org/10.7554/eLife.06426

Mahoney, R., Regan, C., Katona, C., & Livingston, G. (2005). Anxiety and Depression in Family Caregivers of People With Alzheimer Disease: The LASER-AD Study. *The American Journal of Geriatric Psychiatry*, *13*(9), 795–801. https://doi.org/10.1097/00019442-200509000-00008

Maidenbaum, S., Miller, J., Stein, J. M., & Jacobs, J. (2018). Grid-like hexadirectional modulation of human entorhinal theta oscillations. *Proceedings of the National Academy of Sciences*, *115*(42), 10798–10803. https://doi.org/10.1073/pnas.1805007115

Mallory, C. S., Hardcastle, K., Campbell, M. G., Attinger, A., Low, I. I. C., Raymond, J. L., & Giocomo, L. M. (2021). Mouse entorhinal cortex encodes a diverse repertoire of self-motion signals. *Nature Communications*, *12*(1), 671. https://doi.org/10.1038/s41467-021-20936-8

Manly, J. J., & Deters, K. D. (2023). Donanemab for Alzheimer Disease—Who Benefits and Who Is Harmed? *JAMA*, *330*(6), 510–511. https://doi.org/10.1001/jama.2023.11704

Manns, J. R., Stark, C. E., & Squire, L. R. (2000). The visual paired-comparison task as a measure of declarative memory. *Proceedings of the National Academy of Sciences of the United States of America*, *97*(22), 12375–12379. https://doi.org/10.1073/pnas.220398097

Manor, B. R., & Gordon, E. (2003). Defining the temporal threshold for ocular fixation in free-viewing visuocognitive tasks. *Journal of Neuroscience Methods*, *128*(1), 85–93. https://doi.org/10.1016/S0165-0270(03)00151-1

Mäntylä, T., & Holm, L. (2006). Gaze control and recollective experience in face recognition. *Visual Cognition*, *14*(3), 365–386. https://doi.org/10.1080/13506280500347992

Maris, E., & Oostenveld, R. (2007). Nonparametric statistical testing of EEG- and MEG-data. *Journal of Neuroscience Methods*, *164*(1), 177–190. https://doi.org/10.1016/j.jneumeth.2007.03.024

Markostamou, I., & Coventry, K. R. (2022). Age effects on processing spatial relations within different reference frames: The role of executive functions. *Applied Neuropsychology: Adult*, *0*(0), 1–17. https://doi.org/10.1080/23279095.2022.2121212

Martin, A., Brouwers, P., Cox, C., & Fedio, P. (1985). On the nature of the verbal memory deficit in Alzheimer's disease. *Brain and Language*, *25*(2), 323–341. https://doi.org/10.1016/0093-934X(85)90088-4

Mathot, S., Cristino, F., Gilchrist, I. D., & Theeuwes, J. (2012). A simple way to estimate similarity between pairs of eye movement sequences. *Journal of Eye Movement Research*, *5*(1), 1–15. https://doi.org/10.16910/jemr.5.1.4

Mathuranath, P. S., Nestor, P. J., Berrios, G. E., Rakowicz, W., & Hodges, J. R. (2000). A brief cognitive test battery to differentiate Alzheimer's disease and frontotemporal dementia. *Neurology*, *55*(11), 1613–1620. https://doi.org/10.1212/01.wnl.0000434309.85312.19

Matin, E. (1974). Saccadic suppression: A review and an analysis. *Psychological Bulletin*, *81*(12), 899–917. https://doi.org/10.1037/h0037368

Mattsson, N., Zetterberg, H., & Blennow, K. (2010). Lessons from Multicenter Studies on CSF Biomarkers for Alzheimer's Disease. *International Jour-*

nal of Alzheimer's Disease, *2010*, e610613. https://doi.org/10.4061/2010/610613

May, K., Lee, M., Jefford, M., Ribeiro, A., Macdonald, A., Morgan, V., Usher, M., & de Souza, N. M. (2020). Imaging in clinical trials: A patient-led questionnaire study to assess impact of imaging regimes on patient participation. *Research Involvement and Engagement*, *6*(1), 15. https://doi.org/10.1186/s40900-020-00195-5

McGowan, J. W., Kowler, E., Sharma, A., & Chubb, C. (1998). Saccadic localization of random dot targets. *Vision Research*, *38*(6), 895–909. https://doi.org/10.1016/S0042-6989(97)00232-0

McMahon, P. M., Araki, S. S., Sandberg, E. A., Neumann, P. J., & Gazelle, G. S. (2003). Cost-Effectiveness of PET in the Diagnosis of Alzheimer Disease. *Radiology*, *228*(2), 515–522. https://doi.org/10.1148/radiol.2282020915

Mehmood, A., Yang, S., Feng, Z., Wang, M., Ahmad, A. S., Khan, R., Maqsood, M., & Yaqub, M. (2021). A Transfer Learning Approach for Early Diagnosis of Alzheimer's Disease on MRI Images. *Neuroscience*, *460*, 43–52. https://doi.org/10.1016/j.neuroscience.2021.01.002

Mehta, D., Jackson, R., Paul, G., Shi, J., & Sabbagh, M. (2017). Why do trials for Alzheimer's disease drugs keep failing? A discontinued drug perspective for 2010-2015. *Expert Opinion on Investigational Drugs*, *26*(6), 735–739. https://doi.org/10.1080/13543784.2017.1323868

Miller, J. F., Neufang, M., Solway, A., Brandt, A., Trippel, M., Mader, I., Hefft, S., Merkow, M., Polyn, S. M., Jacobs, J., Kahana, M. J., & Schulze-Bonhage, A. (2013). Neural Activity in Human Hippocampal Formation Reveals the Spatial Context of Retrieved Memories. *Science*, *342*(6162), 1111–1114. https://doi.org/10.1126/science.1244056

Mirman, D., Dixon, J. A., & Magnuson, J. S. (2008). Statistical and computational models of the visual world paradigm: Growth curves and individual

differences. *Journal of Memory and Language*, *59*(4), 475–494. https://doi.org/10.1016/j.jml.2007.11.006

Mokrisova, I., Laczo, J., Andel, R., Gazova, I., Vyhnalek, M., Nedelska, Z., Levcik, D., Cerman, J., Vlcek, K., & Hort, J. (2016). Real-space path integration is impaired in Alzheimer's disease and mild cognitive impairment. *Behavioural Brain Research*, *307*, 150–158. https://doi.org/10.1016/j.bbr.2016.03.052

Molitor, R. J., Ko, P. C., & Ally, B. A. (2015). Eye Movements in Alzheimer's Disease. *Journal of Alzheimer's Disease*, *44*(1), 1–12. https://doi.org/10.3233/JAD-141173

Moodley, K., Minati, L., Contarino, V., Prioni, S., Wood, R., Cooper, R., D'Incerti, L., Tagliavini, F., & Chan, D. (2015). Diagnostic differentiation of mild cognitive impairment due to Alzheimer's disease using a hippocampus-dependent test of spatial memory. *Hippocampus*, *25*(8), 939–951. https://doi.org/10.1002/hipo.22417

Mori, E., Yoneda, Y., Yamashita, H., Hirono, N., Ikeda, M., & Yamadori, A. (1997). Medial temporal structures relate to memory impairment in Alzheimer's disease: An MRI volumetric study. *Journal of Neurology, Neurosurgery & Psychiatry*, *63*(2), 214–221. https://doi.org/10.1136/jnnp.63.2.214

Mou, W., & McNamara, T. P. (2002). Intrinsic frames of reference in spatial memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *28*(1), 162–170. https://doi.org/10.1037/0278-7393.28.1.162

Moulin, C. J., James, N., Perfect, T. J., & Jones, R. W. (2003). Knowing What You Cannot Recognise: Further Evidence for Intact Metacognition in Alzheimer's Disease. *Aging, Neuropsychology, and Cognition*, *10*(1), 74–82. https://doi.org/10.1076/anec.10.1.74.13456

Muffato, V., Hilton, C., Meneghetti, C., De Beni, R., & Wiener, J. M. (2019). Evidence for age-related deficits in object-location binding during place

recognition. *Hippocampus*, *29*(10), 971–979. https://doi.org/10.1002/hipo.23099

Myszczynska, M. A., Ojamies, P. N., Lacoste, A. M. B., Neil, D., Saffari, A., Mead, R., Hautbergue, G. M., Holbrook, J. D., & Ferraiuolo, L. (2020). Applications of machine learning to diagnosis and treatment of neurodegenerative diseases. *Nature Reviews Neurology*, *16*(8), 440–456. https://doi.org/10.1038/s41582-020-0377-8

Najemnik, J., & Geisler, W. S. (2005). Optimal eye movement strategies in visual search. *Nature*, *434*(7031), 387–391. https://doi.org/10.1038/nature03390

Nárai, Á., Amora, K. K., Vidnyánszky, Z., & Weiss, B. (2021). Predicting Reading Speed from Eye-Movement Measures. In A. Del Bimbo, R. Cucchiara, S. Sclaroff, G. M. Farinella, T. Mei, M. Bertini, H. J. Escalante, & R. Vezzani (Eds.), *Pattern Recognition. ICPR International Workshops and Challenges* (pp. 453–466). Springer International Publishing. https://doi.org/10.1007/978-3-030-68796-0_33

Nasreddine, Z. S., Phillips, N. A., Bédirian, V., Charbonneau, S., Whitehead, V., Collin, I., Cummings, J. L., & Chertkow, H. (2005). The Montreal Cognitive Assessment, MoCA: A Brief Screening Tool For Mild Cognitive Impairment. *Journal of the American Geriatrics Society*, *53*(4), 695–699. https://doi.org/10.1111/j.1532-5415.2005.53221.x

Nau, M., Julian, J. B., & Doeller, C. F. (2018). How the Brain's Navigation System Shapes Our Visual Experience. *Trends in Cognitive Sciences*, *22*(9), 810–825. https://doi.org/10.1016/j.tics.2018.06.008

Nau, M., Navarro Schröder, T., Bellmund, J. L. S., & Doeller, C. F. (2018). Hexadirectional coding of visual space in human entorhinal cortex. *Nature Neuroscience*, *21*(2), 188–190. https://doi.org/10.1038/s41593-017-0050-8

Negen, J., Heywood-Everett, E., Roome, H. E., & Nardini, M. (2018). Development of allocentric spatial recall from new viewpoints in virtual reality. *Developmental Science*, *21*(1), e12496. https://doi.org/10.1111/desc.12496

Nelson, H. E., & McKenna, P. (1975). The Use of Current Reading Ability in the Assessment of Dementia. *British Journal of Social and Clinical Psychology*, *14*(3), 259–267. https://doi.org/10.1111/j.2044-8260.1975.tb00178.x

Nelson, H. E., & Willison, J. (1991). *The National Adult Reading Test (NART)*. Nfer-Nelson.

Ng, A. Y. (2004). Feature selection, L1 vs. L2 regularization, and rotational invariance. *Proceedings of the twenty-first international conference on Machine learning*, 78. https://doi.org/10.1145/1015330.1015435

Ng, A. Y., Jordan, M. I., & Weiss, Y. (2001). On spectral clustering: Analysis and an algorithm. *Proceedings of the 14th International Conference on Neural Information Processing Systems: Natural and Synthetic*, 849–856.

NICE. (2018, June). Dementia: Assessment, management and support for people living with dementia and their carers. Retrieved November 18, 2023, from https://www.nice.org.uk/guidance/ng97

Nichols, E., Steinmetz, J. D., Vollset, S. E., Fukutaki, K., Chalek, J., Abd-Allah, F., Abdoli, A., Abualhasan, A., Abu-Gharbieh, E., Akram, T. T., Hamad, H. A., Alahdab, F., Alanezi, F. M., Alipour, V., Almustanyir, S., Amu, H., Ansari, I., Arabloo, J., Ashraf, T., … Vos, T. (2022). Estimation of the global prevalence of dementia in 2019 and forecasted prevalence in 2050: An analysis for the Global Burden of Disease Study 2019. *The Lancet Public Health*, *7*(2), e105–e125. https://doi.org/10.1016/S2468-2667(21)00249-8

Noguchi-Shinohara, M., Koike, M., Morise, H., Kudo, K., Tsuchimine, S., Komatsu, J., Abe, C., Kitagawa, S., Ikeda, Y., & Yamada, M. (2021). MEG activity of the dorsolateral prefrontal cortex during optic flow stimulations

detects mild cognitive impairment due to Alzheimer's disease. *PLOS ONE*, *16*(11), e0259677. https://doi.org/10.1371/journal.pone.0259677

Noton, D., & Stark, L. (1971). Scanpaths in Eye Movements during Pattern Perception. *Science*, *171*(3968), 308–311. https://doi.org/10.1126/science.171.3968.308

Odhiambo Omuya, E., Onyango Okeyo, G., & Waema Kimwele, M. (2021). Feature Selection for Classification using Principal Component Analysis and Information Gain. *Expert Systems with Applications*, *174*, 114765. https://doi.org/10.1016/j.eswa.2021.114765

O'Keefe, J., & Burgess, N. (2005). Dual phase and rate coding in hippocampal place cells: Theoretical significance and relationship to entorhinal grid cells. *Hippocampus*, *15*(7), 853–866. https://doi.org/10.1002/hipo.20115

Oleson, J. J., Cavanaugh, J. E., McMurray, B., & Brown, G. (2017). Detecting time-specific differences between temporal nonlinear curves: Analyzing data from the visual world paradigm. *Statistical Methods in Medical Research*, *26*(6), 2708–2725. https://doi.org/10.1177/0962280215607411

Olsen, R. K., Sebanayagam, V., Lee, Y., Moscovitch, M., Grady, C. L., Rosenbaum, R. S., & Ryan, J. D. (2016). The relationship between eye movements and subsequent recognition: Evidence from individual differences and amnesia. *Cortex*, *85*, 182–193. https://doi.org/10.1016/j.cortex.2016.10.007

Opwonya, J., Doan, D. N. T., Kim, S. G., Kim, J. I., Ku, B., Kim, S., Park, S., & Kim, J. U. (2022). Saccadic Eye Movement in Mild Cognitive Impairment and Alzheimer's Disease: A Systematic Review and Meta-Analysis. *Neuropsychology Review*, *32*(2), 193–227. https://doi.org/10.1007/s11065-021-09495-3

Pai, M.-C., & Lee, C.-C. (2016). The Incidence and Recurrence of Getting Lost in Community-Dwelling People with Alzheimer's Disease: A Two and a

Half-Year Follow-Up. *PLOS ONE*, *11*(5), e0155480. https://doi.org/10.1371/journal.pone.0155480

Paraskevaidi, M., Allsop, D., Karim, S., Martin, F. L., & Crean, S. (2020). Diagnostic Biomarkers for Alzheimer's Disease Using Non-Invasive Specimens. *Journal of Clinical Medicine*, *9*(6), 1673. https://doi.org/10.3390/jcm9061673

Parra, M. A., Granada, J., & Fernández, G. (2022). Memory-driven eye movements prospectively predict dementia in people at risk of Alzheimer's disease. *Alzheimer's & Dementia: Diagnosis, Assessment & Disease Monitoring*, *14*(1), e12386. https://doi.org/10.1002/dad2.12386

Passini, R., Rainville, C., Marchand, N., & Joanette, Y. (1995). Wayfinding in dementia of the Alzheimer type: Planning abilities. *Journal of Clinical and Experimental Neuropsychology*, *17*(6), 820–832. https://doi.org/10.1080/01688639508402431

Patterson, C. (2018, September). *World Alzheimer report 2018* (Report). Alzheimer's Disease International. Retrieved November 24, 2023, from https://apo.org.au/node/260056

Pavisic, I. M., Firth, N. C., Parsons, S., Rego, D. M., Shakespeare, T. J., Yong, K. X. X., Slattery, C. F., Paterson, R. W., Foulkes, A. J. M., Macpherson, K., Carton, A. M., Alexander, D. C., Shawe-Taylor, J., Fox, N. C., Schott, J. M., Crutch, S. J., & Primativo, S. (2017). Eyetracking Metrics in Young Onset Alzheimer's Disease: A Window into Cognitive Visual Functions. *Frontiers in Neurology*, *8*. https://doi.org/10.3389/fneur.2017.00377

Pavisic, I. M., Pertzov, Y., Nicholas, J. M., O'Connor, A., Lu, K., Yong, K. X. X., Husain, M., Fox, N. C., & Crutch, S. J. (2021). Eye-tracking indices of impaired encoding of visual short-term memory in familial Alzheimer's disease. *Scientific Reports*, *11*(1), 8696. https://doi.org/10.1038/s41598-021-88001-4

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., & Cournapeau, D. (2011). Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, *12*, 2825–2830.

Pereira, F. R., George, N., Barba, G. D., Dubois, B., & La Corte, V. (2023). The Memory Binding Test: A new tool to assess emerging episodic memory decline in asymptomatic at-risk individuals to Alzheimer's disease. *Alzheimer's & Dementia*, *19*(S18), e074630. https://doi.org/10.1002/alz.074630

Pertzov, Y., Miller, T. D., Gorgoraptis, N., Caine, D., Schott, J. M., Butler, C., & Husain, M. (2013). Binding deficits in memory following medial temporal lobe damage in patients with voltage-gated potassium channel complex antibody-associated limbic encephalitis. *Brain*, *136*(8), 2474–2485. https://doi.org/10.1093/brain/awt129

Petersen, R. C. (2004). Mild cognitive impairment as a diagnostic entity. *Journal of Internal Medicine*, *256*(3), 183–194. https://doi.org/10.1111/j.1365-2796.2004.01388.x

Petersen, R. C., Stevens, J. C., Ganguli, M., Tangalos, E. G., Cummings, J. L., & DeKosky, S. T. (2001). Practice parameter: Early detection of dementia: Mild cognitive impairment (an evidence-based review) [RETIRED]: Report of the Quality Standards Subcommittee of the American Academy of Neurology. *Neurology*, *56*(9), 1133–1142. https://doi.org/10.1212/WNL.56.9.1133

Petersen, R. C. (2009). Early Diagnosis of Alzheimer's Disease: Is MCI Too Late? *Current Alzheimer Research*, *6*(4), 324–330. https://doi.org/10.2174/156720509788929237

Petersen, R. C., Doody, R., Kurz, A., Mohs, R. C., Morris, J. C., Rabins, P. V., Ritchie, K., Rossor, M., Thal, L., & Winblad, B. (2001). Current Concepts

in Mild Cognitive Impairment. *Archives of Neurology*, *58*(12), 1985–1992. https://doi.org/10.1001/archneur.58.12.1985

Petersen, R. C., Smith, G. E., Waring, S. C., Ivnik, R. J., Tangalos, E. G., & Kokmen, E. (1999). Mild Cognitive Impairment: Clinical Characterization and Outcome. *Archives of Neurology*, *56*(3), 303–308. https://doi.org/10.1001/archneur.56.3.303

Pfefferbaum, A., Adalsteinsson, E., & Sullivan, E. V. (2005). Frontal circuitry degradation marks healthy adult aging: Evidence from diffusion tensor imaging. *NeuroImage*, *26*(3), 891–899. https://doi.org/10.1016/j.neuroimage.2005.02.034

Piaget, J., & Inhelder, B. (1956). *The child's conception of space* (F. Langdon & J. Lunzer, Trans.). Routledge.

Ploner, C. J., Gaymard, B. M., Rivaud-Péchoux, S., Baulac, M., Clémenceau, S., Samson, S., & Pierrot-Deseilligny, C. (2000). Lesions Affecting the Parahippocampal Cortex Yield Spatial Memory Deficits in Humans. *Cerebral Cortex*, *10*(12), 1211–1216. https://doi.org/10.1093/cercor/10.12.1211

PostgreSQL. (2022, November). https://www.postgresql.org/

Postma, A., & De Haan, E. H. (1996). What Was Where? Memory for Object Locations. *The Quarterly Journal of Experimental Psychology Section A*, *49*(1), 178–199. https://doi.org/10.1080/713755605

Povova, J., Ambroz, P., Bar, M., Pavukova, V., Sery, O., Tomaskova, H., & Janout, V. (2012). Epidemiological of and risk factors for Alzheimer's disease: A review. *Biomedical Papers*, *156*(2), 108–114. https://doi.org/10.5507/bp.2012.055

Price, J. L., & Morris, J. C. (1999). Tangles and plaques in nondemented aging and "preclinical" Alzheimer's disease. *Annals of Neurology*, *45*(3), 358–368. https://doi.org/10.1002/1531-8249(199903)45:3⟨358::AID-ANA12⟩3.0.CO;2-X

Qi, X., Zhou, Q., Dong, J., & Bao, W. (2023). Noninvasive automatic detection of Alzheimer's disease from spontaneous speech: A review. *Frontiers in Aging Neuroscience*, *15*, 1224723. https://doi.org/10.3389/fnagi.2023.1224723

Ratcliff, R., Thapar, A., & McKoon, G. (2011). Effects of aging and IQ on item and associative memory. *Journal of Experimental Psychology: General*, *140*(3), 464–487. https://doi.org/10.1037/a0023810

Readman, M. R., Polden, M., Gibbs, M. C., Wareing, L., & Crawford, T. J. (2021). The Potential of Naturalistic Eye Movement Tasks in the Diagnosis of Alzheimer's Disease: A Review. *Brain Sciences*, *11*(11), 1503. https://doi.org/10.3390/brainsci11111503

Reagh, Z. M., Noche, J. A., Tustison, N. J., Delisle, D., Murray, E. A., & Yassa, M. A. (2018). Functional Imbalance of Anterolateral Entorhinal Cortex and Hippocampal Dentate/CA3 Underlies Age-Related Object Pattern Separation Deficits. *Neuron*, *97*(5), 1187–1198.e4. https://doi.org/10.1016/j.neuron.2018.01.039

Reitan, R. M. (1958). Validity of the Trail Making Test as an Indicator of Organic Brain Damage. *Perceptual and Motor Skills*, *8*, 271–276. https://doi.org/https://doi.org/10.2466/pms.1958.8.3.271

Rösler, A., Mapstone, M., Hays-Wicklund, A., Gitelman, D. R., & Weintraub, S. (2005). The "Zoom Lens" of Focal Attention in Visual Search: Changes in Aging and Alzheimer's Disease. *Cortex*, *41*(4), 512–519. https://doi.org/10.1016/S0010-9452(08)70191-6

Ruggiero, G., Iavarone, A., & Iachini, T. (2018). Allocentric to Egocentric Spatial Switching: Impairment in aMCI and Alzheimer's Disease Patients? *Current Alzheimer Research*, *15*(3), 229–236. https://doi.org/10.2174/1567205014666171030114821

Ruggiero, G., Ruotolo, F., & Iachini, T. (2012). Egocentric/allocentric and coordinate/categorical haptic encoding in blind people. *Cognitive Processing*, *13*(1), 313–317. https://doi.org/10.1007/s10339-012-0504-6

Ruggiero, G., Ruotolo, F., Iavarone, A., & Iachini, T. (2020). Allocentric coordinate spatial representations are impaired in aMCI and Alzheimer's disease patients. *Behavioural Brain Research*, *393*, 112793. https://doi.org/10.1016/j.bbr.2020.112793

Ryals, A. J., Wang, J. X., Polnaszek, K. L., & Voss, J. L. (2015). Hippocampal contribution to implicit configuration memory expressed via eye movements during scene exploration. *Hippocampus*, *25*(9), 1028–1041. https://doi.org/10.1002/hipo.22425

Ryan, J. D., Shen, K., & Liu, Z.-X. (2020). The intersection between the oculomotor and hippocampal memory systems: Empirical developments and clinical implications. *Annals of the New York Academy of Sciences*, *1464*(1), 115–141. https://doi.org/10.1111/nyas.14256

Salthouse, T. A. (1992). Influence of processing speed on adult age differences in working memory. *Acta Psychologica*, *79*(2), 155–170. https://doi.org/10.1016/0001-6918(92)90030-H

Salvucci, D. D., & Goldberg, J. H. (2000). Identifying fixations and saccades in eye-tracking protocols. *Proceedings of the 2000 symposium on Eye tracking research & applications*, 71–78. https://doi.org/10.1145/355017.355028

Saunders, N. L. J., & Summers, M. J. (2011). Longitudinal deficits to attention, executive, and working memory in subtypes of mild cognitive impairment. *Neuropsychology*, *25*(2), 237–248. https://doi.org/10.1037/a0021134

Schott, J. M., Fox, N. C., Frost, C., Scahill, R. I., Janssen, J. C., Chan, D., Jenkins, R., & Rossor, M. N. (2003). Assessing the onset of structural change in familial Alzheimer's disease. *Annals of Neurology*, *53*(2), 181–188. https://doi.org/10.1002/ana.10424

Scialfa, C. T., Cordazzo, S., Bubric, K., & Lyon, J. (2013). Aging and Visual Crowding. *The Journals of Gerontology: Series B*, *68*(4), 522–528. https://doi.org/10.1093/geronb/gbs086

Seabold, S., & Perktold, J. (2010). Statsmodels: Econometric and Statistical Modeling with Python, 92–96. https://doi.org/10.25080/Majora-92bf1922-011

Segen, V., Avraamides, M. N., Slattery, T. J., & Wiener, J. M. (2020). Age-related differences in visual encoding and response strategies contribute to spatial memory deficits. *Memory & Cognition*. https://doi.org/10.3758/s13421-020-01089-3

Segen, V., Avraamides, M. N., Slattery, T. J., & Wiener, J. M. (2021). Age-related differences in visual encoding and response strategies contribute to spatial memory deficits. *Memory & Cognition*, *49*(2), 249–264. https://doi.org/10.3758/s13421-020-01089-3

Seixas, F. L., Zadrozny, B., Laks, J., Conci, A., & Muchaluat Saade, D. C. (2014). A Bayesian network decision model for supporting the diagnosis of dementia, Alzheimers disease and mild cognitive impairment. *Computers in Biology and Medicine*, *51*, 140–158. https://doi.org/10.1016/j.compbiomed.2014.04.010

Selkoe, D. J. (2011). Resolving controversies on the path to Alzheimer's therapeutics. *Nature Medicine*, *17*(9), 1060–1065. https://doi.org/10.1038/nm.2460

Serino, S., Cipresso, P., Morganti, F., & Riva, G. (2014). The role of egocentric and allocentric abilities in Alzheimer's disease: A systematic review. *Ageing Research Reviews*, *16*, 32–44. https://doi.org/10.1016/j.arr.2014.04.004

Serino, S., Morganti, F., Di Stefano, F., & Riva, G. (2015). Detecting early egocentric and allocentric impairments deficits in Alzheimer's disease: An experimental study with virtual reality. *Frontiers in Aging Neuroscience*,

7. Retrieved October 5, 2023, from https://www.frontiersin.org/articles/10.3389/fnagi.2015.00088

Serino, S., & Riva, G. (2013). Getting lost in Alzheimer's disease: A break in the mental frame syncing. *Medical Hypotheses*, *80*(4), 416–421. https://doi.org/10.1016/j.mehy.2012.12.031

Shakespeare, T. J., Kaski, D., Yong, K. X. X., Paterson, R. W., Slattery, C. F., Ryan, N. S., Schott, J. M., & Crutch, S. J. (2015). Abnormalities of fixation, saccade and pursuit in posterior cortical atrophy. *Brain*, *138*(7), 1976–1991. https://doi.org/10.1093/brain/awv103

Shakespeare, T. J., Pertzov, Y., Yong, K. X. X., Nicholas, J., & Crutch, S. J. (2015). Reduced modulation of scanpaths in response to task demands in posterior cortical atrophy. *Neuropsychologia*, *68*, 190–200. https://doi.org/10.1016/j.neuropsychologia.2015.01.020

Shapiro, S. S., & Wilk, M. B. (1965). An Analysis of Variance Test for Normality (Complete Samples). *Biometrika*, *52*(3/4), 591–611. https://doi.org/10.2307/2333709

Sherrill, K. R., Chrastil, E. R., Ross, R. S., Erdem, U. M., Hasselmo, M. E., & Stern, C. E. (2015). Functional connections between optic flow areas and navigationally responsive brain regions during goal-directed navigation. *NeuroImage*, *118*, 386–396. https://doi.org/10.1016/j.neuroimage.2015.06.009

Shine, J. P., Valdés-Herrera, J. P., Hegarty, M., & Wolbers, T. (2016). The Human Retrosplenial Cortex and Thalamus Code Head Direction in a Global Reference Frame. *Journal of Neuroscience*, *36*(24), 6371–6381. https://doi.org/10.1523/JNEUROSCI.1268-15.2016

Shing, Y. L., Werkle-Bergner, M., Li, S.-C., & Lindenberger, U. (2009). Committing memory errors with high confidence: Older adults do but children don't. *Memory*, *17*(2), 169–179. https://doi.org/10.1080/09658210802190596

Shrager, Y., Kirwan, C. B., & Squire, L. R. (2008). Neural basis of the cognitive map: Path integration does not require hippocampus or entorhinal cortex. *Proceedings of the National Academy of Sciences*, *105*(33), 12034–12038. https://doi.org/10.1073/pnas.0805414105

Simons, D. J., & Wang, R. F. (1998). Perceiving Real-World Viewpoint Changes. *Psychological Science*, *9*(4), 315–320. https://doi.org/https://doi.org/10.1111/1467-9280.00062

Sims, J. R., Zimmer, J. A., Evans, C. D., Lu, M., Ardayfio, P., Sparks, J., Wessels, A. M., Shcherbinin, S., Wang, H., Monkul Nery, E. S., Collins, E. C., Solomon, P., Salloway, S., Apostolova, L. G., Hansson, O., Ritchie, C., Brooks, D. A., Mintun, M., Skovronsky, D. M., & TRAILBLAZER-ALZ 2 Investigators. (2023). Donanemab in Early Symptomatic Alzheimer Disease: The TRAILBLAZER-ALZ 2 Randomized Clinical Trial. *JAMA*, *330*(6), 512–527. https://doi.org/10.1001/jama.2023.13239

Sipatchin, A., Wahl, S., & Rifai, K. (2021). Eye-Tracking for Clinical Ophthalmology with Virtual Reality (VR): A Case Study of the HTC Vive Pro Eye's Usability. *Healthcare*, *9*(2), 180. https://doi.org/10.3390/healthcare9020180

Souchay, C., Isingrini, M., & Gil, R. (2002). Alzheimer's disease and feeling-of-knowing in episodic memory. *Neuropsychologia*, *40*(13), 2386–2396. https://doi.org/10.1016/S0028-3932(02)00075-1

Spieler, D. H., Mayr, U., & Lagrone, S. (2006). Outsourcing cognitive control to the environment: Adult age differences in the use of task cues. *Psychonomic Bulletin & Review*, *13*(5), 787–793. https://doi.org/10.3758/BF03193998

Spiers, H., Burgess, N., Hartley, T., Vargha-Khadem, F., & O'Keefe, J. (2001). Bilateral hippocampal pathology impairs topographical and episodic memory but not visual pattern matching. *Hippocampus*, *11*(6), 715–725. https://doi.org/10.1002/hipo.1087

Sriram, H., Conati, C., & Field, T. (2023). Classification of Alzheimer's Disease with Deep Learning on Eye-tracking Data. *Proceedings of the 25th International Conference on Multimodal Interaction*, 104–113. https://doi.org/10.1145/3577190.3614149

Starr, J. M., & Lonie, J. (2008). Estimated pre-morbid IQ effects on cognitive and functional outcomes in Alzheimer disease: A longitudinal study in a treated cohort. *BMC Psychiatry*, *8*(1), 27. https://doi.org/10.1186/1471-244X-8-27

Sun, J., Liu, Y., Wu, H., Jing, P., & Ji, Y. (2022). A novel deep learning approach for diagnosing Alzheimer's disease based on eye-tracking data. *Frontiers in Human Neuroscience*, *16*. Retrieved November 21, 2023, from https://www.frontiersin.org/articles/10.3389/fnhum.2022.972773

Swainson, R., Hodges, J., Galton, C., Semple, J., Michael, A., Dunn, B., Iddon, J., Robbins, T., & Sahakian, B. (2001). Early Detection and Differential Diagnosis of Alzheimer's Disease and Depression with Neuropsychological Tasks. *Dementia and Geriatric Cognitive Disorders*, *12*(4), 265–280. https://doi.org/10.1159/000051269

Tascón, L., Castillo, J., León, I., & Cimadevilla, J. M. (2018). Walking and non-walking space in an equivalent virtual reality task: Sexual dimorphism and aging decline of spatial abilities. *Behavioural Brain Research*, *347*, 201–208. https://doi.org/10.1016/j.bbr.2018.03.022

Taube, J. S. (1998). Head direction cells and the neurophysiological basis for a sense of direction. *Progress in Neurobiology*, *55*(3), 225–256. https://doi.org/10.1016/S0301-0082(98)00004-5

Taube, J. S. (2007). The Head Direction Signal: Origins and Sensory-Motor Integration. *Annual Review of Neuroscience*, *30*(1), 181–207. https://doi.org/10.1146/annurev.neuro.29.051605.112854

Tavares, G., Perona, P., & Rangel, A. (2017). The Attentional Drift Diffusion Model of Simple Perceptual Decision-Making. *Frontiers in Neuroscience*,

11. Retrieved November 23, 2023, from https://www.frontiersin.org/articles/10.3389/fnins.2017.00468

Thal, D. R., Rüb, U., Orantes, M., & Braak, H. (2002). Phases of A-beta-deposition in the human brain and its relevance for the development of AD. *Neurology*, *58*(12), 1791–1800. https://doi.org/10.1212/WNL.58.12.1791

Thomas, A. W., Molter, F., Krajbich, I., Heekeren, H. R., & Mohr, P. N. C. (2019). Gaze bias differences capture individual choice behaviour. *Nature Human Behaviour*, *3*(6), 625–635. https://doi.org/10.1038/s41562-019-0584-8

Tromp, J., Klotzsche, F., Krohn, S., Akbal, M., Pohl, L., Quinque, E. M., Belger, J., Villringer, A., & Gaebler, M. (2020, September). *OpenVirtualObjects (OVO): An open set of standardized and validated 3D household objects for virtual reality-based research, assessment, and therapy.* (preprint). Open Science Framework. Retrieved October 1, 2020, from https://osf.io/ejkzt

Tsatali, M., Poptsi, E., Moraitou, D., Agogiatou, C., Bakoglidou, E., Gialaouzidis, M., Papasozomenou, C., Soumpourou, A., & Tsolaki, M. (2021). Discriminant Validity of the WAIS-R Digit Symbol Substitution Test in Subjective Cognitive Decline, Mild Cognitive Impairment (Amnestic Subtype) and Alzheimer's Disease Dementia (ADD) in Greece. *Brain Sciences*, *11*(7), 881. https://doi.org/10.3390/brainsci11070881

Tu, M.-C., & Pai, M.-C. (2006). Getting lost for the first time in patients with Alzheimer's disease. *International Psychogeriatrics*, *18*(3), 567–570. https://doi.org/10.1017/S1041610206224025

Vallat, R. (2018). Pingouin: Statistics in Python. *Journal of Open Source Software*, *3*(31), 1026. https://doi.org/10.21105/joss.01026

Valliappan, N., Dai, N., Steinberg, E., He, J., Rogers, K., Ramachandran, V., Xu, P., Shojaeizadeh, M., Guo, L., Kohlhoff, K., & Navalpakkam, V. (2020).

Accelerating eye movement research via accurate and affordable smartphone eye tracking. *Nature Communications*, *11*(1), 4553. https://doi.org/10.1038/s41467-020-18360-5

Valve. (2020). Half-Life: Alyx.

van der Ham, I. J. M., Claessen, M. H. G., Evers, A. W. M., & van der Kuil, M. N. A. (2020). Large-scale assessment of human navigation ability across the lifespan. *Scientific Reports*, *10*(1), 3299. https://doi.org/10.1038/s41598-020-60302-0

Varesi, A., Carrara, A., Pires, V. G., Floris, V., Pierella, E., Savioli, G., Prasad, S., Esposito, C., Ricevuti, G., Chirumbolo, S., & Pascale, A. (2022). Blood-Based Biomarkers for Alzheimer's Disease Diagnosis and Progression: An Overview. *Cells*, *11*(8), 1367. https://doi.org/10.3390/cells11081367

Veneri, G., & Rufa, A. (2017). Extrafoveal Vision Maximizes the Likelihood to Grab Information in Visual-sequential Search. *Computer Communication & Collaboration*, *5*(1), 21–34.

Venini, D., Remington, R. W., Horstmann, G., & Becker, S. I. (2014). Centre-of-Gravity Fixations in Visual Search: When Looking at Nothing Helps to Find Something. *Journal of Ophthalmology*, *2014*, e237812. https://doi.org/10.1155/2014/237812

Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., van der Walt, S. J., Brett, M., Wilson, J., Millman, K. J., Mayorov, N., Nelson, A. R. J., Jones, E., Kern, R., Larson, E., ... van Mulbregt, P. (2020). SciPy 1.0: Fundamental algorithms for scientific computing in Python. *Nature Methods*, *17*(3), 261–272. https://doi.org/10.1038/s41592-019-0686-2

Vishwanath, D., & Kowler, E. (2003). Localization of shapes: Eye movements and perception compared. *Vision Research*, *43*(15), 1637–1653. https://doi.org/10.1016/S0042-6989(03)00168-8

Vitek, G. E., Decourt, B., & Sabbagh, M. N. (2023). Lecanemab (BAN2401): An anti–beta-amyloid monoclonal antibody for the treatment of Alzheimer disease. *Expert Opinion on Investigational Drugs*, *32*(2), 89–94. https://doi.org/10.1080/13543784.2023.2178414

Wang, D., Mulvey, F. B., Pelz, J. B., & Holmqvist, K. (2017). A study of artificial eyes for the measurement of precision in eye-trackers. *Behavior Research Methods*, *49*(3), 947–959. https://doi.org/10.3758/s13428-016-0755-8

Wang, J., Guo, X., Zhuang, X., Chen, T., & Yan, W. (2017). Disrupted pursuit compensation during self-motion perception in early Alzheimer's disease. *Scientific Reports*, *7*(1), 4049. https://doi.org/10.1038/s41598-017-04377-2

Wang, K. N., Page, A. T., & Etherton-Beer, C. D. (2021). Mild cognitive impairment: To diagnose or not to diagnose. *Australasian Journal on Ageing*, *40*(2), 111–115. https://doi.org/10.1111/ajag.12913

Wang, R. F., & Simons, D. J. (1999). Active and passive scene recognition across views. *Cognition*, *70*(2), 191–210. https://doi.org/10.1016/S0010-0277(99)00012-8

Ward, J. H. (1963). Hierarchical Grouping to Optimize an Objective Function. *Journal of the American Statistical Association*, *58*(301), 236–244. https://doi.org/10.1080/01621459.1963.10500845

Wechsler, D. (2019, February). Wechsler Adult Intelligence Scale-Third Edition. https://doi.org/10.1037/t49755-000

Weniger, G., Ruhleder, M., Lange, C., Wolf, S., & Irle, E. (2011). Egocentric and allocentric memory as assessed by virtual reality in individuals with amnestic mild cognitive impairment. *Neuropsychologia*, *49*(3), 518–527. https://doi.org/10.1016/j.neuropsychologia.2010.12.031

Whitehouse, P. J., & George, D. R. (2016). A Tale of Two Reports: What Recent Publications from the Alzheimer's Association and Institute of Medicine

say about the State of the Field. *Journal of Alzheimer's Disease*, *49*(1), 21–25. https://doi.org/10.3233/JAD-150663

WHO. (2023, March). Dementia fact sheet. https://www.who.int/news-room/fact-sheets/detail/dementia

Wilcockson, T. D., Mardanbegi, D., Xia, B., Taylor, S., Sawyer, P., Gellersen, H. W., Leroi, I., Killick, R., & Crawford, T. J. (2019). Abnormalities of saccadic eye movements in dementia due to Alzheimer's disease and mild cognitive impairment. *Aging (Albany NY)*, *11*(15), 5389–5398. https://doi.org/10.18632/aging.102118

Winblad, B., Amouyel, P., Andrieu, S., Ballard, C., Brayne, C., Brodaty, H., Cedazo-Minguez, A., Dubois, B., Edvardsson, D., Feldman, H., Fratiglioni, L., Frisoni, G. B., Gauthier, S., Georges, J., Graff, C., Iqbal, K., Jessen, F., Johansson, G., Jönsson, L., ... Zetterberg, H. (2016). Defeating Alzheimer's disease and other dementias: A priority for European science and society. *The Lancet Neurology*, *15*(5), 455–532. https://doi.org/10.1016/S1474-4422(16)00062-4

Wisse, L. E. M., Butala, N., Das, S. R., Davatzikos, C., Dickerson, B. C., Vaishnavi, S. N., Yushkevich, P. A., & Wolk, D. A. (2015). Suspected non-AD pathology in mild cognitive impairment. *Neurobiology of Aging*, *36*(12), 3152–3162. https://doi.org/10.1016/j.neurobiolaging.2015.08.029

Wittenberg, R., Knapp, M., Karagiannidou, M., Dickson, J., & Schott, J. M. (2019). Economic impacts of introducing diagnostics for mild cognitive impairment Alzheimer's disease patients. *Alzheimer's & Dementia: Translational Research & Clinical Interventions*, *5*, 382–387. https://doi.org/10.1016/j.trci.2019.06.001

Wolbers, T., & Hegarty, M. (2010). What determines our navigational abilities? *Trends in Cognitive Sciences*, *14*(3), 138–146. https://doi.org/10.1016/j.tics.2010.01.001

Wold, S., Esbensen, K., & Geladi, P. (1987). Principal component analysis. *Chemometrics and Intelligent Laboratory Systems*, *2*(1), 37–52. https://doi.org/10.1016/0169-7439(87)80084-9

Wolfsgruber, S., Polcher, A., Koppara, A., Kleineidam, L., Frölich, L., Peters, O., Hüll, M., Rüther, E., Wiltfang, J., Maier, W., Kornhuber, J., Lewczuk, P., Jessen, F., & Wagner, M. (2017). Cerebrospinal Fluid Biomarkers and Clinical Progression in Patients with Subjective Cognitive Decline and Mild Cognitive Impairment. *Journal of Alzheimer's Disease*, *58*(3), 939–950. https://doi.org/10.3233/JAD-161252

Wollstadt, P., Hasenjäger, M., & Wiebel-Herboth, C. B. (2021). Quantifying the Predictability of Visual Scanpaths Using Active Information Storage. *Entropy*, *23*(2), 167. https://doi.org/10.3390/e23020167

Wynn, J. S., Bone, M. B., Dragan, M. C., Hoffman, K. L., Buchsbaum, B. R., & Ryan, J. D. (2016). Selective scanpath repetition during memory-guided visual search. *Visual Cognition*, *24*(1), 15–37. https://doi.org/10.1080/13506285.2016.1175531

Wynn, J. S., Liu, Z.-X., & Ryan, J. D. (2022). Neural Correlates of Subsequent Memory-Related Gaze Reinstatement. *Journal of Cognitive Neuroscience*, *34*(9), 1547–1562. https://doi.org/10.1162/jocn_a_01761

Wynn, J. S., Olsen, R. K., Binns, M. A., Buchsbaum, B. R., & Ryan, J. D. (2018). Fixation reinstatement supports visuospatial memory in older adults. *Journal of Experimental Psychology: Human Perception and Performance*, *44*(7), 1119–1127. https://doi.org/10.1037/xhp0000522

Wynn, J. S., Shen, K., & Ryan, J. D. (2019). Eye Movements Actively Reinstate Spatiotemporal Mnemonic Content. *Vision*, *3*(2), 21. https://doi.org/10.3390/vision3020021

Yamanakkanavar, N., Choi, J. Y., & Lee, B. (2020). MRI Segmentation and Classification of Human Brain Using Deep Learning for Diagnosis of

Alzheimer's Disease: A Survey. *Sensors*, *20*(11), 3243. https://doi.org/10.3390/s20113243

Yarbus, A. L. (1967). Eye Movements During Fixation on Stationary Objects. In A. L. Yarbus (Ed.), *Eye Movements and Vision* (pp. 103–127). Springer US. Retrieved December 1, 2023, from https://doi.org/10.1007/978-1-4899-5379-7_4

Yeung, L.-K., Hale, C., Rizvi, B., Igwe, K., Sloan, R. P., Honig, L. S., Small, S. A., & Brickman, A. M. (2020). Anterolateral Entorhinal Cortex Volume is Associated with Memory Retention in Clinically Unimpaired Older Adults. *Neurobiology of Aging*. https://doi.org/10.1016/j.neurobiolaging.2020.10.031

Yeung, L.-K., Olsen, R. K., Hong, B., Mihajlovic, V., D'Angelo, M. C., Kacollja, A., Ryan, J. D., & Barense, M. D. (2019). Object-in-place Memory Predicted by Anterolateral Entorhinal Cortex and Parahippocampal Cortex Volume in Older Adults. *Journal of Cognitive Neuroscience*, *31*(5), 711–729. https://doi.org/10.1162/jocn_a_01385

Ying, J., Reboreda, A., Yoshida, M., & Brandon, M. P. (2023). Grid cell disruption in a mouse model of early Alzheimer's disease reflects reduced integration of self-motion cues. *Current Biology*, *33*(12), 2425–2437.e5. https://doi.org/10.1016/j.cub.2023.04.065

Yu, H., Wang, K., Zhong, P., Cheng, H.-D., Lv, X.-Y., & Yuan, L.-L. (2020). Investigations of Memory Monitoring in Individuals With Subjective Cognitive Decline and Amnestic Mild Cognitive Impairment. *Cognitive and Behavioral Neurology*, *33*(3), 201. https://doi.org/10.1097/WNN.0000000000000242

Zaccara, G., Gangemi, P. F., Muscas, G. C., Paganini, M., Pallanti, S., Parigi, A., Messori, A., & Arnetoli, G. (1992). Smooth-pursuit eye movements: Alterations in Alzheimer's disease. *Journal of the Neurological Sciences*, *112*(1), 81–89. https://doi.org/10.1016/0022-510X(92)90136-9

Zelinsky, G. J. (2012). TAM: Explaining off-object fixations and central fixation tendencies as effects of population averaging during search. *Visual Cognition*, *20*(4-5), 515–545. https://doi.org/10.1080/13506285.2012.666577

Zelinsky, G. J., & Loschky, L. C. (2005). Eye movements serialize memory for objects in scenes. *Perception & Psychophysics*, *67*(4), 676–690. https://doi.org/10.3758/BF03193524

Zelinsky, G. J., Rao, R. P. N., Hayhoe, M. M., & Ballard, D. H. (1997). Eye Movements Reveal the Spatiotemporal Dynamics of Visual Search. *Psychological Science*, *8*(6), 448–453. https://doi.org/10.1111/j.1467-9280.1997.tb00459.x

Zemblys, R., Niehorster, D. C., & Holmqvist, K. (2019). gazeNet: End-to-end eye-movement event detection with deep neural networks. *Behavior Research Methods*, *51*(2), 840–864. https://doi.org/10.3758/s13428-018-1133-5

Zhang, F., Li, Z., Zhang, B., Du, H., Wang, B., & Zhang, X. (2019). Multi-modal deep learning model for auxiliary diagnosis of Alzheimer's disease. *Neurocomputing*, *361*, 185–195. https://doi.org/10.1016/j.neucom.2019.04.093

Zhang, J.-X., Wang, L., Hou, H.-Y., Yue, C.-L., Wang, L., & Li, H.-J. (2021). Age-related impairment of navigation and strategy in virtual star maze. *BMC Geriatrics*, *21*(1), 108. https://doi.org/10.1186/s12877-021-02034-y

Zola, S. M., Squire, L. R., Teng, E., Stefanacci, L., Buffalo, E. A., & Clark, R. E. (2000). Impaired recognition memory in monkeys after damage limited to the hippocampal region. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *20*(1), 451–463. https://doi.org/10.1523/JNEUROSCI.20-01-00451.2000

Zola, S. M., Manzanares, C. M., Clopton, P., Lah, J. J., & Levey, A. I. (2013). A Behavioral Task Predicts Conversion to Mild Cognitive Impairment and

Alzheimer's Disease. *American Journal of Alzheimer's Disease & Other Dementias®*, *28*(2), 179–184. https://doi.org/https://doi.org/10.1177/1533317512470484